AWARD NUMBER: W81XWH-14-1-0596

TITLE: Biomarkers for Early Detection of Clinically Relevant Prostate Cancer: A Multi-Institutional Validation Trial

PRINCIPAL INVESTIGATOR: Jesse McKenney, MD

CONTRACTING ORGANIZATION: Cleveland Clinic Foundation
Cleveland, OH 44195

REPORT DATE: October 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
October 2017	Annual	30 Sept 2016 – 29 Sept. 2017

4. TITLE AND SUBTITLE
Biomarkers for Early Detection of Clinically Relevant Prostate Cancer:
A Multi-Institutional Validation Trial

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0596

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jesse McKenney, MD

E-Mail: MCKENNJ@CCF.ORG

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cleveland Clinic Foundation
9500 Euclid Ave.
Cleveland, OH 44195

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
For men diagnosed with early stage prostate cancer a critical need exists for molecular assays that accurately distinguish aggressive prostate cancer from those cancers that will not cause harm if left untreated. In this project, we are assessing three different panels of established molecular biomarkers for their ability to distinguish aggressive cancers from indolent cancers. We have established agreements with three commercial companies to analyze their biomarker platforms in our multi-center, prospectively accrued prostate cancer active surveillance cohort – the Canary Prostate Active Surveillance Study (PASS). We are in the process of evaluating these three biomarker panels in tissue, blood, and urine samples with well annotated clinical and pathologic data collected as part of PASS. We are conducting rigorous statistical evaluation to demonstrate the utility and performance of biomarkers in clinical practice to predict aggressive disease. The accuracy of each biomarker for predicting short- and long-term progression will be characterized with time dependent receiver operating characteristic curves. The successful clinical validation of biomarkers that offer substantially improved predictive and prognostic accuracy should bring extraordinary potential to improve the care of prostate cancer patients.

15. SUBJECT TERMS
Prostate cancer, active surveillance, biomarkers, validation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	96	USAMRMC

Table of Contents

1. INTRODUCTION

Although prostate-specific antigen (PSA) testing and the resulting treatment of prostate cancer (PCa) is likely responsible for some of the 44% decrease in prostate cancer mortality witnessed in the United States since 1992, the detection of low risk tumors has increased. The majority of prostate cancers currently diagnosed are low risk tumors for which there is substantial evidence that the cancer will not cause harm if left untreated. However, enough uncertainty remains in accurately identifying which tumors will not cause harm to a patient that many low risk cancers are still treated, resulting in so-called overtreatment. To reduce this overtreatment, while still diagnosing aggressive high risk tumors early enough that they can be successfully treated, there is a critical need for molecular assays that accurately distinguish more aggressive disease from cancers that will not cause harm. The goal of this project is to perform rigorous clinical validation of established biomarkers in order to improve the accuracy of risk assessment and distinguish aggressive from indolent disease in men with apparently low-risk disease by standard clinical variables. We are evaluating multiple established and analytically validated quantitative molecular biomarkers to predict PCa progression in a multi-center active surveillance cohort with high-quality biospecimens. We aim to unlink the diagnosis of PCa with immediate treatment, thus addressing the overtreatment issue and economic, physical, and emotional burdens of PCa diagnoses. The results have promise to change the standard of care in the treatment of the majority of newly diagnosed PCa with near term impact due to the availability of the biomarkers and execution in an established, prospective cohort of men undergoing AS.

2. KEYWORDS

Prostate cancer; active surveillance; progression; aggressive disease; central pathology review; biomarkers; prediction models; PCA3; TMPRSS2:ERG; kallikreins; 4Kscore; OncotypeDX;

3. ACCOMPLISHMENTS

What were the major goals and objectives of the project?

We hypothesize that biomarkers of disease aggressiveness and prognosis can be interrogated in low risk prostate cancer (PCa) and that these biomarkers will better detect clinically relevant PCa in asymptomatic patients, thus distinguishing aggressive from indolent disease and immediately impacting both the initial choice of therapy and decision-making during AS. The objective of the study is to utilize analytically validated assays that take into account tumor heterogeneity to measure biomarkers in specimens that were collected in a non-invasive manner.

The major goals of the project, as stated in the scope of work, are:

1. Collection of specimens and clinical data. (Coordinated by FHCRC)

Milestone 1. Completion of a minimum of three years of follow-up with high-quality data and specimen collection. Due: 12/30/2016 **COMPLETED**

2. Analysis of scientific aim 1: Validate a panel of tissue-based biomarkers to determine the presence of or progression to aggressive disease. (Lead site: FHCRC)

Milestone 2. Execute collaboration agreement with GHI. Due 12/30/2014 **COMPLETED**.
Milestone 3. Tissue blocks identified for analysis. Due: 12/30/2015 **COMPLETED**
Milestone 4. Oncotype DX validation complete in PASS cohort. Due 12/30/2016 *in process*
Milestone 5. Manuscript submission of Oncotype DX validation. Due 9/30/2017 *in process*

3. Analysis of scientific aim 2: Evaluate a panel of four-kallikrein plasma-based markers to determine the presence of or progression to clinically relevant prostate cancer. (Lead site: FHCRC)

Milestone 6. Execute collaboration agreement with OPKO. Due 3/30/2015 **COMPLETED**.
Milestone 7. Plasma samples identified for analysis. Due 12/30/2015 **COMPLETED**
Milestone 8. OPKO 4KScore validation complete in PASS cohort. Due 9/30/2016 **COMPLETED**
Milestone 9. Manuscript submission of 4KScore validation. Due 9/30/2017 **COMPLETED**

4. Analysis of scientific Aim 3: Confirm the ability of PCA3 mRNA concentrations in urine, alone or in combination with TMPRSS2:ERG mRNA. (Lead site: FHCRC)

Milestone 10. Urine specimens identified for analysis. Due 12/30/2014 **COMPLETED**
Milestone 11. PCA3 and TMPRSS2:ERG validation complete in PASS cohort. Due 12/30/2015 **COMPLETED**
Milestone 12. Manuscript submission of PCA3 and TMPRSS2:ERG validation. Due 9/30/2017 **COMPLETED**

5. Central pathology review of PASS biopsy and RP slides. (Lead site: CCF)
   Milestone 13. Completion of Central Pathology Review for biopsy-driven endpoints. Due: 12/30/2016 *in process*

6. Translation of biomarkers into clinical practice. (Lead sites: FHCRC and CCF)
   Milestone 14. Construction of integrated model of biomarkers for the prediction of progression in the PASS cohort. Due 9/30/2017 *in process*
   Milestone 15. Manuscript submission of integrated model for prediction of progression. Due 9/30/2017 *in process*

What was accomplished under these goals?

Task 1: Collection of specimens and clinical data. (Coordinated by FHCRC)

Collection of follow-up data and longitudinal specimens in the PASS cohort is essential to adequately power our funded biomarker analyses. To date, PASS has enrolled 1,534 eligible patients at nine clinical sites. Outside of the scope of this proposal, Emory has been added as a new PASS site and they have enrolled 30 participants to date. We have been highly successful in following participants to obtain outcomes measures, with a median cohort follow-up of over 4.2 years (25th and 75th percentiles: 2.2, 6.0 years). Currently, all of the first 1000 participants enrolled in PASS, which are the subject of this specific research proposal, have at least three years of follow-up. In the past year, we have conducted site visits to two clinical sites (University of British Columbia on 10/18/2016 and Eastern Virginia Medical School on 7/19/2017) to ensure adherence to the protocol. The coordinating center based at the Fred Hutchinson Cancer Center continues to provide data QA and QC.

Task 2: Analysis of scientific Aim 1: Validate a panel of tissue-based biomarkers to determine the presence of or progression to aggressive disease. (Lead site: FHCRC)

We have completed collection and delivery of the FFPE tissue blocks from diagnostic biopsies of 634 PASS participants to Genomic Health, Inc. (GHI). To obtain the tissue required reconsenting PASS participants for use of the tissue left over from initial diagnosis, and collection of the tissue blocks from pathology departments both at the PASS sites and at local urology clinics. All blocks were sent to the PASS Central Biospecimen Repository where they underwent QC and were labeled with unique PASS ID numbers prior to sending to GHI. See **Table 1** for distribution of the specimens collected by site. GHI sectioned all tissue blocks to make 8 unstained sections (unless the sectioning was performed at the PASS site's pathology department, in which case unstained slides were delivered to GHI.) The top and bottom section were stained with H&E. Dr. McKenney at Cleveland Clinic reviewed all H&E slides provided to him by GHI and has recorded the tumor extent and Gleason Score, as described in the GHI protocol. GHI macro dissects the tumor tissue and isolates RNA for the Oncotype DX

assay, and finally they calculate a GPS score. Of the 634 cases, 185 (29%) have been identified as pathology failures, meaning that there is insufficient tumor tissue for Oncotype DX assay. GHI is currently processing the final batches of specimens and expects to provide the resulting GPS scores to the FHCRC analysts no later than December 2017.

Table 1. Collection of Diagnostic Tissue for Oncotype DX Assay

Clinical Site	**No. of Cases Sent to GHI**	**No. of RP Cases**
UTHSCSA	100	28
UW	186	51
BIDMC	91	14
Stanford	49	8
UBC	84	23
VA-Seattle	54	13
U Mich	25	2
EVMS	45	7
Total	**634**	**146**

Task 3: Analysis of scientific Aim 2: Evaluate a panel of four-kallikrein plasma-based markers to determine the presence of or progression to clinically relevant prostate cancer. (Lead site: FHCRC)

We have collaborated with OPKO to assay a panel of four kallikreins (total PSA (tPSA), free PSA (fPSA), intact PSA (iPSA), and human kallikrein 2 (hK2)). Statistical models were developed to predict reclassification from Gleason 6 cancer to Gleason 7 or greater. The analysis plan was determined before specimens were selected for the study, and included breaking the data/specimens into training and testing cohorts, using a 2/3 to 1/3 split. The models included clinical information and either the 4Kpanel or serum PSA. We used Receiver Operating Characteristic (ROC) curve analyses and area under the curve (AUC) to assess discriminatory capacity and decision curve analysis (DCA) to report clinical net benefit.

Significant predictors for reclassification were 4Kpanel (OR=1.54 [1.31,1.81]) or PSA (OR=2.11 [1.53,2.91]), ≥20% cores positive (OR=2.10 [1.33,3.32]), ≥2 prior negative biopsies (OR=0.19 [0.04,0.85]), prostate volume (OR=0.47 [0.31,0.70]), BMI (OR=1.09 [1.04,1.14]). ROC curve analysis comparing 4Kpanel and base models indicated that the 4Kpanel improved accuracy for predicting reclassification (AUC 0.78 versus 0.74) in the first surveillance biopsy. Both models performed comparably for prediction of reclassification in subsequent biopsies (AUC=0.75 versus 0.76). In DCA, both models showed higher net benefit compared to biopsy-all and biopsy-none strategies.

Conclusions: The 4Kpanel provided incremental value over routine clinical information in predicting high-grade cancer in the first biopsy after diagnosis. The 4Kpanel did not add predictive value to the base model at subsequent surveillance biopsies.

These results were presented at the 2016 Meeting of the American Urological Association (AUA) and have been published in European Urology (v72, pp448-454.) A reprint of the final publication is included with this report.

Task 4: Analysis of Specific Aim 3: Confirm the ability of PCA3 mRNA concentrations in urine, alone or in combination with TMPRSS2:ERG mRNA, to predict the presence of or development to clinically relevant prostate cancer. (Lead site: FHCRC)

PCA3 and the TMPRSS2:ERG fusion are prostate cancer-specific biomarkers that hold promise for stratifying risk in the setting of AS. Hologic Gen-Probe's assay to quantitate urine PCA3 transcripts in post-digital rectal exam (DRE) urine is FDA-approved for men with a previous negative biopsy, given peer reviewed evidence that it can reduce unnecessary prostate biopsies. We aim to confirm the ability of the PCA3 and TMPRSS2:ERG assays to predict aggressive prostate cancer in the entire PASS cohort. To this end, we have collaborated with Hologic Gen-Probe to analyze 2,069 urine specimens from 783 PASS participants. We have submitted a manuscript to Clinical Cancer Research for publication. A copy of the manuscript is included with this report in Appendix 2.

Task 5: Central Pathology Review (Lead Site: CCF)

The purpose of central pathology review is to standardize endpoints for analyses of biomarkers. With funding from this grant, we have developed a customized pathology review system, in which primary and secondary pathology reviewers can access scanned images and record key data from each slide. All data recorded by the primary and secondary reviewers are reviewed for consistency and if results are discrepant, a consensus review is conducted to resolve. In an early analysis of scoring, we evaluated slides from 131 unique diagnostic biopsies, collected from five different PASS study sites. In this small subset, 71% of cases were reviewed concordantly by study pathologists and the original pathologist (**Table 2**). The 29% discordant reviews highlight the need for a centralized review of cases to obtain accurate data, as Gleason is used as an endpoint in many biomarker studies.

We have made good progress on this aim but have not yet completed the scanning of slides. We have scanned 1577 slides from 956 biopsies. All biopsies are reviewed by the primary pathologist as well as a secondary reviewer.

We are also collecting the H&E slides from radical prostatectomies (RP) performed after a period of active surveillance on PASS participants. To date there are about 250 RPs, and Dr. Jesse McKenney, the PASS Central Pathologist, has reviewed about 110.

Table 2. Analysis of Concordance in Central Pathology Review of 131 PASS Biopsies

Original Path		Total Scenarios				
Gleason	Total Cases	Total Agreement	CR Agreement	Orig & 1° Agree	Orig & 2° Agree	Total Disagreement
3 + 3	120	90 (76)	12 (10)	11 (9)	4 (3)	3 (2)
3 + 4	8	2 (25)	1 (13)	0	5 (63)	0
4 + 3	3	0	3 (67)	0	0	0
TOTAL	**131**	**92 (71)**	**16 (12)**	**11 (9)**	**9 (7)**	**3 (2)**

Task 6. Translation of biomarkers into clinical practice. (Lead sites: FHCRC and CCF)

To use any biomarker in clinical practice we feel that it is essential that the biomarkers add incremental improvement in prediction of high grade or high volume disease over the clinical variables alone. Until very recently there have been no risk prediction models for use in active surveillance, and the ones that have recently been developed have not fully utilized important variables available in contemporary clinical practice. Thus, as an important component of translating biomarkers into clinical management of active surveillance patients, we have been developing base risk prediction models using commonly available clinical variables (PSA, prostate size, and biopsy information from the diagnostic and surveillance biopsies). These models are used as the basis for evaluation of biomarkers for clinical management, but they may by themselves be exceedingly useful in managing active surveillance patients.

To utilize risk prediction models in clinical management, we are developing user interfaces that not only calculate the risk of finding aggressive cancer (e.g. high grade cancer) at the next or future biopsies, but also show important metrics of how well a given model may be predicting risk and what the clinical consequences are of using a specific risk threshold in decision making. Our first such risk prediction tool uses the “base” model developed to evaluate the 4 Kallikreins (Lin et. al., European Urology, v72, pp448-454, 2017). In this model, a patient’s age, body mass index, prostate size, PSA, and information about the ratio of prior biopsy cores containing cancer and of prior biopsies in which no cancer was found are used to predict the risk of finding high grade (Gleason ≥7) cancer in the next biopsy. In addition to the risk of finding high grade cancer, the calculator displays the 95% confidence interval of that risk, how the risk compares to the active surveillance (PASS) population, and where a given patient’s value for each variable falls within the population from which the model was developed (**Figure 1**). The clinical consequences of using a specific risk threshold to guide the decision of whether or not to perform a biopsy are shown on a separate tab of the calculator

(**Figure 2**). A patient report, or a simpler depiction of the risk information, is also available in the calculator (**Figure 3**). The PASS Risk Calculator can currently be found at: https://canarypass.shinyapps.io/biopsy_nomogram/

Figure 1. First tab of PASS Risk Calculator.

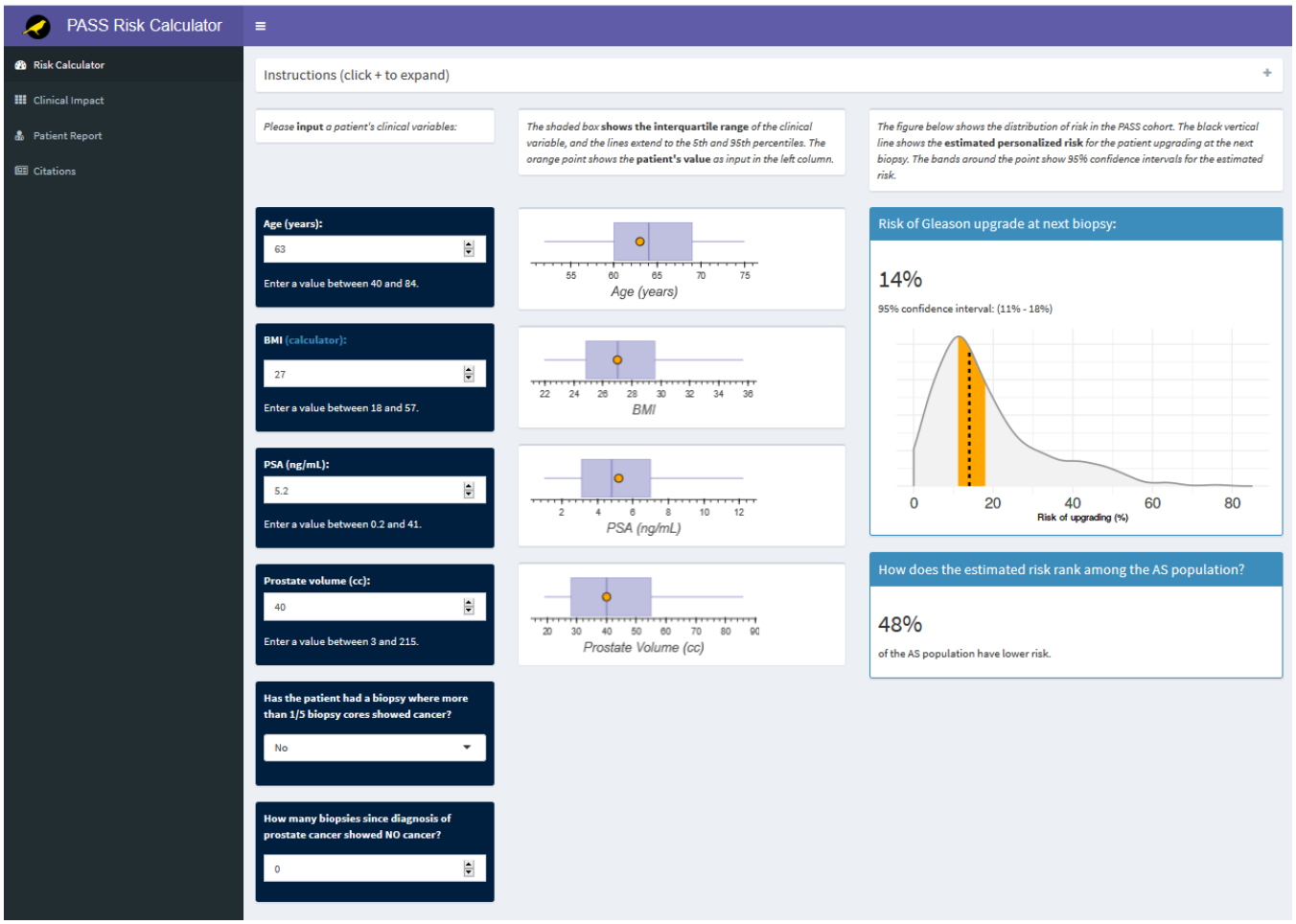


Figure 2. Clinical Impact tab of PASS Risk Calculator.

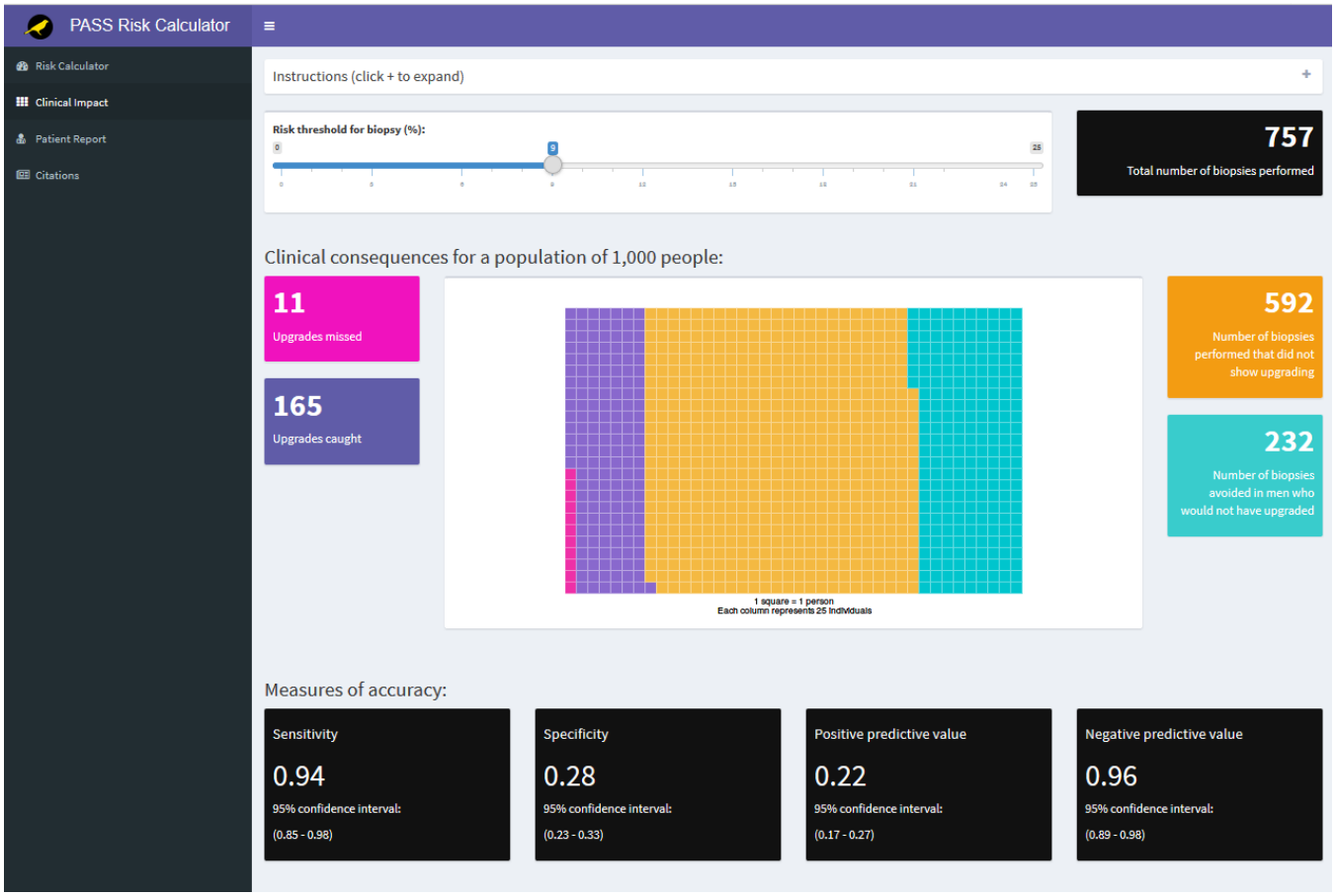
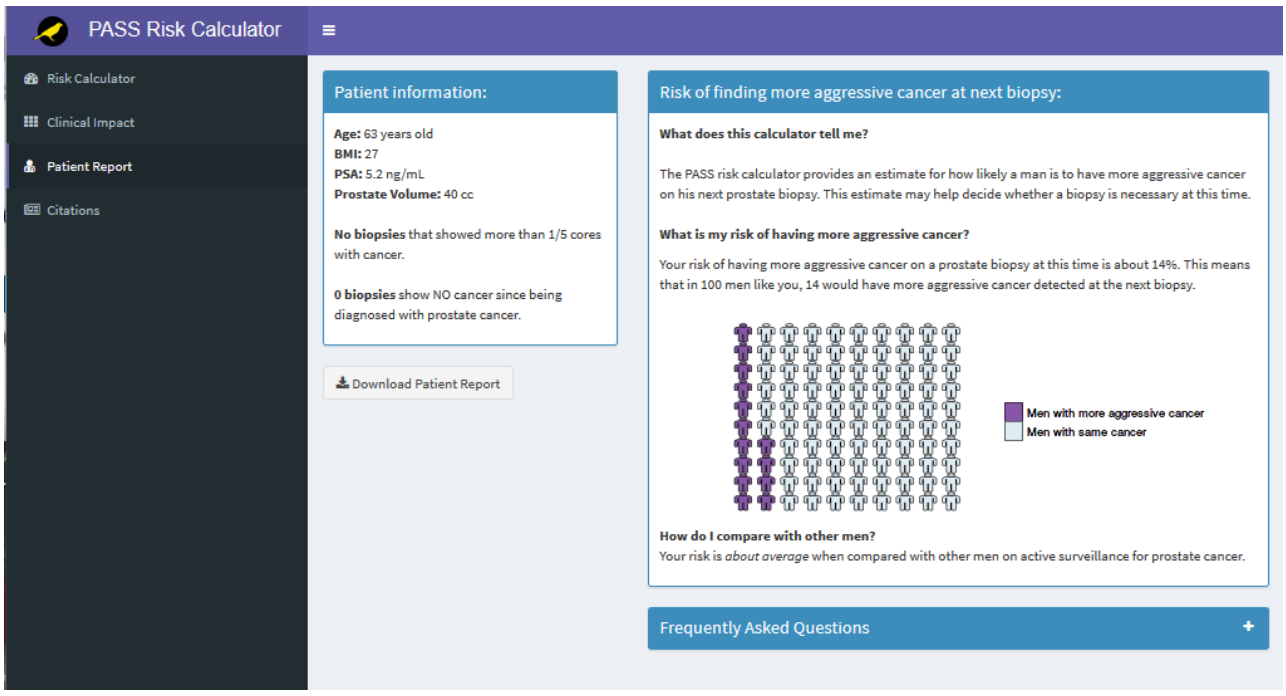


Figure 3. Patient Report of PASS Risk Calculator.



To improve upon our risk prediction models, we have been evaluating individual parameters. We have found that PSA kinetics, when calculated using a linear mixed effects model, is associated with time to biopsy reclassification in a model adjusting for prostate size, time since diagnosis, biopsy parameters, and PSA at diagnosis. The details of this work are described in the attached paper titled “Refined analysis of prostate specific antigen kinetics to predict prostate cancer active surveillance outcomes.” We have also found that after a diagnosis of low grade prostate cancer, having one or two biopsies in which no cancer are found substantially reduces a man’s risk of upgrading at future biopsies. The details of this work are described in the attached manuscript titled “The Role of Surveillance Biopsy with No Cancer as a Prognostic Marker for Reclassification: Results from the Canary Prostate Active Surveillance Study (PASS).”

What opportunities for training and professional development did the project provide?
Nothing to report. This grant does not provide for training or professional development activities.

How were the results disseminated to communities of interest?
Results are being disseminated through presentations at national meetings and through publication.

What do you plan to do during the next reporting period to accomplish the goals and objectives?
Our plans for the next year of funding are as follows:

- We expect Oncotype DX assay data and GPS scores from GHI by December, 2017. These data will be merged with PASS data (Task 2e in SOW).
- Statistical analysis will be conducted on the merged dataset (Task 2f).
- Reporting and manuscript preparation of above findings will be completed (Task 2g).
- We will complete central pathology review of remaining biopsies and prostatectomy cases (Task 5 in SOW).
- We will evaluate an integrated model of correlative biomarkers (Task 6).

4**. IMPACT**

What was the impact on the development of the principal discipline(s) of the project?

We anticipate that the successful clinical validation of biomarkers that offer substantially improved predictive and prognostic accuracy would bring extraordinary potential to improve the care of PCa patients. Specifically, those men with clinically low-risk tumors that can be confirmed as truly low-risk with greater accuracy could be spared the cost and quality-of-life impact of invasive diagnostic and therapeutic procedures. Conversely, those men with apparent low-risk disease who in fact harbor higher-risk tumors or have the potential to develop lethal disease will be identified, thus avoiding under-treatment. Such a paradigm shift in PCa care would yield near-term changes in the PCa treatment landscape, greatly improving the cost-benefit calculations for population-level PCa screening efforts and reducing the overtreatment of disease.

What was the impact on other disciplines?

Nothing to report in this period, although we expect that statistical techniques being developed will be utilized to evaluate biomarker performance in many diseases other than prostate cancer.

What was the impact on technology transfer?

This project involves evaluation and validation of commercial biomarker panels that have not previously been used in the active surveillance setting. While we do not expect a direct impact on technology transfer, there should be a large impact on the commercial use of the molecular diagnostics.

What was the impact on society beyond science and technology?

Successful execution of this project should transform the clinical management of prostate cancer in several ways. First, if patients and their physicians have a reliable and valid estimate of the risks of disease progression and harm, then more might opt for surveillance, thereby reducing the risks of overtreatment and its attendant substantial costs and morbidity. Such improved accuracy would allow men to be selected more appropriately and with greater confidence for surveillance rather than immediate treatment. Second, a proportion of men initially choosing active surveillance eventually opt for primary curative treatment even with no objective measures of clinical progression, presumably due to patient/provider anxiety. Increasing patient and provider confidence in risk assessments would presumably lead to increased adherence to active surveillance, further decreasing overtreatment. Third, a marker panel with high accuracy for progression on active surveillance will influence the regimen of clinical re-assessment, such that those men with particularly low-risk disease might be eligible

for a less intensive surveillance protocol with fewer repeated prostate biopsies, reducing the use of the most invasive, and risky, component of a typical surveillance regimen. Fourth, the proposed markers might also facilitate treatment planning for men not currently on surveillance. For example, a man with apparently low-risk disease but a significantly adverse biomarker panel would have an increased risk of occult high-grade disease and perhaps should undergo staging lymphadenectomy at time of prostatectomy, a procedure which might not routinely be performed for low risk disease. Lastly, the public health impact of a validated biomarker panel will be substantial, as the costs of initial curative therapy for prostate cancer accounts for $2-3 billion annually. Approximately half of the new diagnoses are low risk cancers and candidates for active surveillance, and accurate determination of who may benefit from curative therapy, while sparing the majority, would have immediate economic impact.

5. CHANGES / PROBLEMS

Changes in approach and reasons for change
Nothing to report.

Actual or anticipated problems or delays and actions or plans to resolve them
We have experienced unanticipated delays in obtaining pathology slides and prostate biopsy tissue blocks to complete three of the milestones in our Statement of Work. Delays were due to slow response-time of local pathology facilities from whom we were requesting samples. Specific milestones impacted are:

- Milestone 4: Oncotype DX validation complete in PASS cohort
- Milestone 5: Manuscript submission of Oncotype DX validation
- Milestone 13: Completion of Central Pathology Review for biopsy-driven endpoints

Changes that had a significant impact on expenditures
With the unanticipated delays in obtaining pathology slides and prostate biopsy tissue blocks, our spending has slowed temporarily. The delay in Milestones 4, 5, and 13 have resulted in a temporary reduction in FTE and a lower than anticipated spending on shipping and scanning costs. While the work is taking longer than anticipated to complete, we still are on track to complete the project in its entirety and will require all of our awarded funds. An extension without funds has been approved for 12 months to complete all milestones.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents:

- **Significant changes in use or care of human subjects:** No significant changes in the use or care of human subjects. The Fred Hutchinson Cancer Research Center has approved the study activities through 5/29/2018 under IR file number 8271. A continuing review was submitted to HRPO (Log Number A-18320) and receipt was acknowledged on 06/14/2017.
- **Significant changes in use or care of vertebrate animals:** Nothing to report.
- **Significant changes in use of biohazards and/or select agents:** Nothing to report.

6. PRODUCTS

Publications, Conference Papers, and Presentations

Journal publications

Daniel W. Lin, Lisa F. Newcomb, Marshall D. Brown, Daniel D. Sjoberg, Yan Dong, James D. Brooks, Peter R. Carroll, Matthew Cooperberg, Atreya Dash, William J. Ellis, Michael Fabrizio, Martin E. Gleave, Todd M. Morgan, Peter S. Nelson, Ian M. Thompson, Andrew Wagner, and Yingye Zheng for the Canary Prostate Active Surveillance Study Investigators. “Evaluating the four kallikrein panel of the 4Kscore for prediction of high-grade prostate cancer in 2 men in the Canary Prostate Active Surveillance Study (PASS).” European Urology.
Acknowledgement of federal support: Yes.

Matthew R. Cooperberg; James D. Brooks; Anna V. Faino; Lisa F. Newcomb; James T. Kearns; Peter R. Carroll; Atreya Dash; Ruth Etzioni; Michael D. Fabrizio; Martin E. Gleave; Todd M. Morgan; Peter S. Nelson; Ian M. Thompson; Andrew A. Wagner; Daniel W. Lin,; and Yingye Zheng. “Refined analysis of prostate specific antigen kinetics to predict prostate cancer active surveillance outcomes.” European Urology. Under review.
Acknowledgement of federal support: Yes.

Lisa F. Newcomb, Yingye Zheng, Anna V. Faino, Daniella Bianchi-Frias, Matthew R. Cooperberg, Marshall D. Brown, James D. Brooks, Peter R. Carroll, Atreya Dash, Michael D. Fabrizio, Martin E. Gleave, Michael Liss, Todd M. Morgan, Ian M. Thompson, Andrew A. Wagner, Peter S. Nelson, and Daniel W. Lin. “Performance of PCA3 and TMPRSS2:ERG urinary biomarkers in prediction of biopsy outcome in the Canary Prostate Active Surveillance Study (PASS).” Clinical Cancer Research. Under review.
Acknowledgement of federal support: Yes.

James T. Kearns, Anna V. Faino, Lisa F. Newcomb, James D. Brooks, Peter R. Carroll, Atreya Dash, William J. Ellis, Michael Fabrizio, Martin E. Gleave, Todd M. Morgan, Peter S. Nelson, Ian M. Thompson, Andrew A. Wagner, Yingye Zheng, and Daniel W. Lin. “The Role of Surveillance Biopsy with No Cancer as a Prognostic Marker for Reclassification: Results from the Canary Prostate Active Surveillance Study (PASS).” European Urology. Under review.
Acknowledgement of federal support: Yes.

Books or other non-periodical, one-time publications

Nothing to report.

Other publications, conference papers, and presentations

Lin D, Brown M, Newcomb L, Sjoberg D, Brooks J, Carroll P, Dash A, Fabrizio M, Gleave M, Morgan T, Nelson P, Thompson I, Zheng Y. PD08-02: “Evaluating the four kallikrein panel of the 4KScore for prediction of high-grade prostate cancer in men in the Canary Prostate Active Surveillance Study (PASS).” Annual Meeting of the American Urological Association; 2016 May 6-10, San Diego, CA.

Newcomb L. “Evaluating urinary PCA3 and TMPRSS2:ERG for prediction of adverse biopsy reclassification in men in the Canary Prostate Active Surveillance Study (PASS).” Presentation at the Multi-Institutional Prostate Cancer SPORE Program Retreat, 2016 March 13-15, Fort Lauderdale, FL.

Website(s) or other Internet site(s)

PASS Risk Calculator can be found at: https://canarypass.shinyapps.io/biopsy_nomogram/ The link to this will be at canarypass.org in the near future.

Technologies or techniques

Nothing to report.

Inventions, patent applications, and/or licenses

Nothing to report.

Other Products

As part of this project we continue to maintain a large biospecimen repository with associated clinical and demographic data, which serves as a rich resource for the scientific community. In the coming years of this award we anticipate scientific results, validated diagnostics, and prediction models that should make an impact on the clinical management of patients with prostate cancer.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	Daniel Lin, MD
Project Role:	Principal Investigator
Researcher Identifier (e.g. ORCID ID):	ORCID: 0000-0002-2135-1534
Nearest person month worked:	2 person months
Contribution to Project:	As Principal Investigator, Dr. Lin oversees the execution of the project, including interactions with industry collaborators and the FDA. He directs overall scientific activities including data collection, interpretation, and manuscript preparation. Dr. Lin takes a central role in the analysis of all data from the project, collaborating with the other investigators on manuscript preparations.
Funding Support:	N/A

Name:	Jesse McKenney, MD
Project Role:	Principal Investigator of Partner Award
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	2 person months
Contribution to Project:	Dr. McKenney is the lead pathologist for this project, overseeing all aspects of the central pathology review. He has worked on development of the Centralized Pathology Review system, and leads the group of study pathologists who review all endpoints for PASS participants. He ensures that pathologic review is timely and follows project guidelines.
Funding Support:	N/A

Name:	Hilary Boyer
Project Role:	Research Scientist
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	3 person months
Contribution to Project:	Ms. Boyer works under the direction of Dr. Newcomb to receive, annotate, and track PASS specimens from the Central Repository. Ms. Boyer is responsible for pulling, tracking, and documenting specimens sent to collaborating sites and coordinates all shipping activities. She also assists in specimen and clinical data QA and

	QC, in monitoring study progress, and in preparing reports for study investigators.
Funding Support:	N/A

Name:	Anna Faino, MS
Project Role:	Statistical Research Associate
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	5 person months
Contribution to Project:	Ms. Faino works under the supervision of Dr. Zheng and is responsible for the extensive data analysis involved in this project. She participates in study consultation with project investigators and the data operations group on data and database forms. Under Dr. Zheng's supervision she performs data analyses, data interpretation and manuscript preparation.
Funding Support:	N/A

Name:	Suzanne Kolb, MPH
Project Role:	Project Coordinator
Researcher Identifier (e.g. ORCID ID):	ORCID: 0000-0002-6443-644X
Nearest person month worked:	5 person months
Contribution to Project:	Ms. Kolb works under the direction of Drs. Lin and Newcomb to fulfill daily fiscal and administrative functions of the program. She monitors subaward budgets and provides logistical support. Ms. Kolb works closely with the PASS Deputy Director to maintain IRB files, material transfer agreements, and other regulatory documents as well as tracking project timelines and deliverables.
Funding Support:	N/A
Name:	Lisa Newcomb, PhD
Project Role:	Deputy Director
Researcher Identifier (e.g. ORCID ID):	ORCID: 0000-0003-3505-3754
Nearest person month worked:	2 person months
Contribution to Project:	Dr. Newcomb facilitates the day-to-day operations of all aspects of the research, interfacing with the PASS Study to ensure high quality data and specimens. She works closely with Dr. Lin and all investigators and collaborators in the execution of the project. Dr. Newcomb is responsible for specimen selection,

	management of the acquisition and distribution of specimens from the biorepository, as well as overseeing regulatory requirements and supervising study staff.
Funding Support:	N/A

Name:	Maria Tretiakova, MD, PhD
Project Role:	Co-investigator, Pathologist
Researcher Identifier (e.g. ORCID ID):	ORCID: 0000-0002-0819-9638
Nearest person month worked:	2 person months
Contribution to Project:	Dr. Tretiakova is responsible for reviewing slides of prostate needle biopsies and characterizing the pathologic parameters such as Gleason score and amount of cancer. She is also working with co-investigators at FHCRC and Cleveland Clinic on study design, data analysis, and interpretation.
Funding Support:	N/A

Name:	Lawrence True, MD
Project Role:	Pathologist
Researcher Identifier (e.g. ORCID ID):	ORCID: 0000-0002-8621-9569
Nearest person month worked:	1 person month
Contribution to Project:	Dr. True is responsible for reviewing slides of prostate needle biopsies and characterizing the pathologic parameters such as Gleason score and amount of cancer. He is working with co-investigators at FHCRC and Cleveland Clinic on study design, data analysis, and interpretation.
Funding Support:	N/A

Name:	William A. Willette
Project Role:	Programmer
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	1 person months
Contribution to Project:	Mr. Willette is responsible for customizing and maintaining the PASS study database as well as the central pathology review system. This includes creation of custom slide views, annotation forms, and reports to facilitate the pathology review workflow and collect and monitor the pathology review data. Mr. Willette also prepares reports for investigators and the PASS team.

Funding Support:	N/A

Name:	Yingye Zheng, PhD
Project Role:	Co-investigator, Biostatistician
Researcher Identifier (e.g. ORCID ID):	ORCID: 0000-0002-3078-4200
Nearest person month worked:	1 person months
Contribution to Project:	Dr. Zheng is responsible for all statistical aspects of this project, including design and analysis. She consults with investigators on study designs and necessary study design modifications if necessary during the course of the study. She ensures that appropriate data items are collected for valid data analyses and QA/QC to be conducted to ensure high quality of clinical and assay data. She also supervises the SRA in data analyses and interpretation of study data.
Funding Support:	N/A

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Yes. Listed below are changes in the other support for senior and key personnel. Please note: none of these changes impacts effort on the project.

LIN, D.

Funding Ended:

U01 CA199338 (Etzioni) 9/1/15 – 8/31/17 0.24 cal months
NIH $22,956
"Modeling to Improve Cancer Outcomes Across Diverse Populations (CISNET)"

MCKENNEY, J.

No changes .

TRETIAKOVA, M.

Funding Ended:

W81XWH-14-2-0183 (Vessella/Morrissey) 9/30/14-9/30/17 0.6 cal months
DOD $540,925
"Prostate cancer biorepository network."

Prostate Cancer Foundation 12/24/14-12/24-16 0.6 cal months
Subaward from FHCRC (Nelson) $400,000

"Eradicating lethal micrometastatic cancer through high intensity short course AR suppression."

TRUE, L.

New Funding:

W81XWH-16-1-0584 (Petros)	9/30/16 – 9/29/19	0.07 cal months
DOD	$18,416	

"A Single Missense Mutation in 77% of Prostate Cancer Bone Metastasis: Novel Opportunity for Genetic Biomarker and Novel Therapeutic Mitochondrial Target."

Funding Ended:

P50 CA97186-11 (Mostaghel)	1/1/16 – 12/31/16	0.12 cal months
NIH/NCI	$27,178	

"Development of an Automated Image Analysis Protocol and Analytic Algorithm to Quantify Relative Levels of AR and AR Variants in Prostate Cancer (Pilot Project)"

PC130652 (Tomlins)	7/1/14 – 6/30/17	.96 cal months
DOD	$105,620	

"Clonal evaluation of prostate cancer by ERG/SPINK1 status to improve prognosis prediction."

ZHENG, Y.

New Funding:

RO1 CA192438 (Malone/Li)	12/1/17 – 11/30/19	0.6 cal months
NIH	$531,451	

"Epidemiology of the four most frequent cancers following breast cancer."

Funded Ended:

PO1 CA053996 (Prentice)	7/1/11 – 6/30/16	1.8 cal months
NIH	$566,278	

"Statistical methods for medical studies (Project 3)"

RO1 (Jinbo Chen)	4/1/13 – 3/30/17	0.6 cal months
NIH	$169,876	

"Statistical methods for the development of absolute risk prediction models."

What other organizations were involved as partners?

Organization Name: University of Washington

Location of Organization: Seattle, WA
Partner's contribution to the project:

Facilities: Staff (Drs. Lin, True, Tretiakova) used facilities provided by the University of Washington for pathology review and office space.

Collaboration: University of Washington personnel provide expertise in pathology (Drs. Tretiakova and True) and study oversight (Dr. Newcomb).

Organization Name: Cleveland Clinic
Location of Organization: Cleveland, OH
Partner's contribution to the project:

Facilities: Dr. McKenney uses facilities provided by the Cleveland Clinic for central pathology review.

Collaboration: Dr. McKenney provides expertise for central pathology review.

Organization Name: Genomic Health, Inc.
Location of Organization: Redwood City, CA
Partner's contribution to the project:

Collaboration: Genomic Health, Inc. has agreed to run Prostate Oncotype Dx assays free of charge and discussed design of project.

Organization Name: OPKO Diagnostics
Location of Organization: Miami, FL
Partner's contribution to the project:

Collaboration: OPKO Diagnostics has run the blood kallikrein assays free of charge and discussed design of project.

Organization Name: Hologic GenProbe
Location of Organization: San Diego, CA
Partner's contribution to the project:

Collaboration: Hologic GenProbe has run the PCA3 and TMPRSS2:ERG urine marker assays free of charge.

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS:
For this project, Dr. Daniel Lin is the initiating PI and Dr. Jesse McKenney is the partnering PI. Drs. Lin and McKenney are independently submitting a duplicate annual project report, with tasks clearly marked with the responsible PI and research site as requested.

QUAD CHARTS: Not applicable.

9. APPENDICES

Appendix 1. Published 4KScore Paper. European Urology 72 (2017) 448-454. Pages 26 – 32.

Appendix 2. Submitted manuscript, titled “Refined analysis of prostate specific antigen kinetics to predict prostate cancer active surveillance outcomes.” Submitted to European Urology. Pages 33 – 55.

Appendix 3. Submitted manuscript, titled “Performance of PCA3 and TMPRSS2:ERG urinary biomarkers in prediction of biopsy outcome in the Canary Prostate Active Surveillance Study (PASS)”. Submitted to Clinical Cancer Research. Pages 56 – 76.

Appendix 4. Submitted manuscript, titled “The role of surveillance biopsy with no cancer as a prognostic marker for reclassification: Results from the Canary Prostate Active Surveillance Study (PASS).” Under review at European Urology. Pages 77 – 96.

available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Prostate Cancer

Evaluating the Four Kallikrein Panel of the 4Kscore for Prediction of High-grade Prostate Cancer in Men in the Canary Prostate Active Surveillance Study

Daniel W. Lin[a,b,c,*], *Lisa F. Newcomb*[a,b], *Marshall D. Brown*[a], *Daniel D. Sjoberg*[d], *Yan Dong*[e], *James D. Brooks*[f], *Peter R. Carroll*[g], *Matthew Cooperberg*[g], *Atreya Dash*[c], *William J. Ellis*[b], *Michael Fabrizio*[h], *Martin E. Gleave*[i], *Todd M. Morgan*[j], *Peter S. Nelson*[a], *Ian M. Thompson*[k], *Andrew A. Wagner*[l], *Yingye Zheng*[a],

for the Canary Prostate Active Surveillance Study Investigators

[a] *Fred Hutchinson Cancer Research Center, Seattle, WA, USA;* [b] *University of Washington, Seattle, WA, USA;* [c] *Veterans Affairs Puget Sound Health Care System, Seattle, WA, USA;* [d] *Memorial Sloan-Kettering Cancer Center, New York, NY, USA;* [e] *OPKO Labs, Nashville, TN, USA;* [f] *Stanford University, Stanford, CA, USA;* [g] *University of California at San Francisco, San Francisco, CA, USA;* [h] *Eastern Virginia Medical School, Norfolk, VA, USA;* [i] *University of British Columbia, Vancouver, Canada;* [j] *University of Michigan, Ann Arbor, MI, USA;* [k] *University of Texas Health Sciences Center at San Antonio, San Antonio, TX, USA;* [l] *Beth Israel Deaconess Medical Center, Boston, MA, USA*

Article info

Article history:
Accepted November 11, 2016

Associate Editor:
Giacomo Novara

Keywords:
Prostatic neoplasms
Prospective studies
Active surveillance
Biomarker
Kallikrein

Abstract

Background: Diagnosis of Gleason 6 prostate cancer can leave uncertainty about the presence of undetected aggressive disease.
Objective: To evaluate the utility of a four kallikrein (4K) panel in predicting the presence of high-grade cancer in men on active surveillance.
Design, setting, and participants: Plasma collected before the first and subsequent surveillance biopsies was assessed for 718 men prospectively enrolled in the multi-institutional Canary PASS trial. Biopsy data were split 2:1 into training and test sets. We developed statistical models that included clinical information and either the 4Kpanel or serum prostate-specific antigen (PSA).
Outcome measurements and statistical analysis: The endpoint was reclassification to Gleason $\geq$7. We used receiver operating characteristic (ROC) curve analyses and area under the curve (AUC) to assess discriminatory capacity, and decision curve analysis (DCA) to report clinical net benefit.
Results and limitations: Significant predictors for reclassification were 4Kpanel (odds ratio [OR] 1.54, 95% confidence interval [CI] 1.31–1.81) or PSA (OR 2.11, 95% CI 1.53–2.91), $\geq$20% cores positive (OR 2.10, 95% CI 1.33–3.32), two or more prior negative biopsies (OR 0.19, 95% CI 0.04–0.85), prostate volume (OR 0.47, 95% CI 0.31–0.70), and body mass index (OR 1.09, 95% CI 1.04–1.14). ROC curve analysis comparing 4K and base models indicated that the 4Kpanel improved accuracy for predicting reclassification (AUC 0.78 vs 0.74) at the first surveillance biopsy. Both models performed comparably for prediction of reclassification at subsequent biopsies (AUC 0.75 vs 0.76). In DCA, both models showed higher net benefit compared to biopsy-all and biopsy-none strategies. Limitations include the single cohort nature of the study and the small numbers; results should be validated in another cohort before clinical use.

* Corresponding author. Department of Urology, University of Washington, 1959 NE Pacific Street, Box 356510, Seattle, WA 98195, USA. Tel. +1 206 2210797; Fax: +1 206 5433272.
E-mail address: dlin@uw.edu (D.W. Lin).

http://dx.doi.org/10.1016/j.eururo.2016.11.017

Conclusions: The 4Kpanel provided incremental value over routine clinical information in predicting high-grade cancer in the first biopsy after diagnosis. The 4Kpanel did not add predictive value to the base model at subsequent surveillance biopsies.
Patient summary: Active surveillance is a management strategy for many low-grade prostate cancers. Repeat biopsies monitor for previously undetected high-grade cancer. We show that a model with clinical variables, including a panel of four kallikreins, indicates the presence of high-grade cancer before a biopsy is performed.

1. Introduction

Active surveillance is a management strategy for low-grade, localized prostate cancer that allows men to delay or be spared the potential morbidities of treatment. Cancers that appear to be low-risk at diagnosis are monitored, typically with serial prostate-specific antigen (PSA) measurements, clinical examinations, and repeat prostate biopsies. Intervention is recommended on evidence of a more aggressive tumor, usually based on changes in biopsy characteristics.

However, fear of occult high-grade cancer, in part because of the known undersampling of systematic prostate biopsies, has tempered widespread adoption of active surveillance. Even with emerging magnetic resonance imaging (MRI)–based biopsy protocols, there remains uncertainty surrounding the presence of more aggressive disease against a background of apparently low-risk cancer. In addition, the optimal surveillance schedule and triggers for intervention have not been established, resulting in substantial variations in the practice of active surveillance. Prostate biopsy can be painful, anxiety-provoking, expensive, and potentially morbid, so avoiding unnecessary surveillance biopsies is attractive. Methods to reduce the number of biopsies in active surveillance regimens, while maximizing the identification of high-grade cancers that may benefit from treatment, would have substantial clinical utility.

A promising approach to determine active surveillance candidacy and surveillance regimens (eg, more intensive vs less intensive biopsy schedules) involves the addition of biomarker panels to prediction models based on known clinical and demographic variables [1]. Among men suspected of having prostate cancer, a panel of four kallikreins (total PSA [tPSA], free PSA [fPSA], intact PSA [iPSA], and human kallikrein 2 [hK2]) combined with age using a mathematical algorithm improves the prediction of high-grade cancers compared to the PCPT risk calculator or models using tPSA alone [2,3]. Here, we explore the utility of prediction models incorporating the predefined four kallikrein panel algorithm (4Kpanel) to predict the presence of occult high-grade disease in men already diagnosed with Gleason 6 cancer and on active surveillance. We use plasma specimens and data from the prospective, multi-institutional Canary Prostate Active Surveillance Study (PASS).

2. Patients and methods

2.1. Study cohort

This study included men from Canary PASS, a multicenter, prospective study enrolling men on active surveillance [4]. Participants in PASS consented to specimen collection as part of the PASS protocol (clinicaltrials.gov NCT00756665), which was approved by institutional review boards at participating sites. The PASS protocol includes monitoring at clinic visits every 6 mo, with the first ≥10-core prostate needle biopsy at 6–12 mo, the second at 24 mo after cancer diagnosis, and subsequent biopsies every 2 yr. Specimens, including EDTA plasma, were collected at study entry and every 6-mo clinic visit, and were stored at −70 °C until use.

In February 2015, 1170 participants were enrolled in PASS at nine sites throughout North America. Of these, 956 participants had an on-study biopsy, of whom 877 had Gleason 3 + 3 disease at study entry, 771 had not used 5α-reductase inhibitors, and EDTA plasma collected before biopsy was available for 753 men. Participants with missing prostate volume or ratio of positive to total biopsy cores were excluded from the modeling ($n = 35$); the remaining 718 men, who had 1111 biopsies, were included in this study.

2.2. Laboratory methods

Blood was collected in K_2EDTA vacutainers, inverted, centrifuged at 1600 × *g*, and frozen at −70 °C within 4 h of collection. Frozen plasma was stored until shipment on dry ice to OPKO Labs (Nashville, TN, USA) for analysis. The analysis laboratory was blinded to all specimen and clinical information. Specimens were thawed immediately before analysis. tPSA, fPSA, iPSA, and hK2 were measured [2].

2.3. Study design and analyses

The objective of the analyses was to determine whether a model using clinical predictors and kallikrein data collected after diagnosis of Gleason 6 cancer, but before surveillance biopsy, can predict high-grade cancer in the surveillance biopsy. Sequential surveillance biopsies were considered as two groups: (1) the initial biopsy after cancer diagnosis (sometimes called confirmatory biopsy) and (2) all subsequent surveillance biopsies. Biopsy data were split 2:1 into training and test sets matched by outcome.

The primary outcome was reclassification from Gleason score 6 to Gleason score ≥7. A value for the 4Kpanel was calculated with tPSA, fPSA, iPSA, hk2, and age using locked down coefficients developed before the study was conducted [3]. This combination of the four kallikreins is the same as in the commercial 4Kscore. However, the commercial 4Kscore is a model containing the 4Kpanel and clinical data available before cancer diagnosis, and is calibrated for a patient before diagnosis. Because we evaluated the kallikreins in a cohort already diagnosed with cancer, we developed a new model that included the 4Kpanel and clinical information available after a diagnosis of cancer, and calibrated to an active surveillance population. Additional clinical predictors considered in modeling included age, body mass index (BMI), race (African American or other), digital rectal examination (DRE) results, number of previous biopsies after diagnosis, number of negative biopsies after diagnosis, core ratio (ratio of biopsy cores containing cancer to total cores) from previous biopsy, maximum core ratio among all previous biopsies, months since diagnosis, and prostate volume (prostate size measured closest to the time of sampling and imputed within 2 yr).

Table 1 – Characteristics for 478 participants with kallikreins assayed before the initial surveillance biopsy after diagnosis for combined Gleason score <7 versus ≥7 for the training and test cohorts

Characteristics	Training set			Test set		
	Gleason <7	Gleason ≥7	*p* value	Gleason <7	Gleason ≥7	*p* value
Sample size (*n*)	259	60		125	34	
Age at diagnosis (yr)	63 (58–67)	64 (60–68)	0.109	64 (58–68)	64 (57–67)	0.876
Body mass index (kg/m^2)	27 (25–30)	28 (25–33)	0.116	27 (25–29)	28 (26–31)	0.305
Race						
Non–African American	248 (96)	56 (93)		121 (97)	29 (85)	
African American	11 (4)	4 (7)	0.646	4 (3)	5 (15)	0.522
Time from diagnosis (mo)	12.0 (8.4–14.1)	12.7 (8.6–14.8)	0.237	12.2 (8.8–14.0)	12.6 (10.3–17.6)	0.189
Digital rectal examination						
Normal	238 (92)	55 (92)		118 (94)	30 (88)	
Abnormal	21 (8)	5 (8)	0.971	7 (6)	4 (12)	0.031
Prostate volume (cm^3)	41.0 (30.0–56.5)	35.5 (25.0–50.0)	0.041	40.0 (30.0–51.0)	30.0 (24.0–42.8)	0.006
Positive:total core ratio	0.08 (0.08–0.17)	0.17 (0.08–0.20)	<0.001	0.08 (0.08–0.17)	0.17 (0.17–0.25)	<0.001
Clinical serum PSA (ng/ml)	4.60 (2.91–6.40)	4.81 (4.35–6.42)	0.108	4.56 (3.11–6.24)	5.65 (4.58–7.88)	0.024
4Kpanel (logit)	0.21 (0.08–0.29)	0.32 (0.16–0.44)	<0.001	0.20 (0.07–0.28)	0.36 (0.18–0.53)	<0.001

PSA = prostate-specific antigen.
Data are presented as median (interquartile range) for continuous variables and as n (%) for categorical variables.

Either the 4Kpanel (logit scale) or clinical serum PSA (log-transformed) was used in models. Prediction models were built using data in the training set, and then clinical performance was assessed using the testing set. We followed the principles set forth by the US Food and Drug Administration critical path initiative, using an established biomarker with analytic validity for the intent of clinical validation in the intended use population [7]. Furthermore, we followed reporting recommendations for tumor marker prognostic studies (REMARK) [8] and the Tumor Marker Utility Grading System [9] in reporting the clinical utility of the biomarker panel.

2.3.1. Model building

Data from initial and subsequent biopsy groups were combined for model development. Interaction terms between biopsy group (initial vs subsequent surveillance biopsy) and other variables were evaluated to investigate whether effects may differ for an initial biopsy and a subsequent biopsy. Logistic regression was used to fit the models, with robust variance to account for the correlation among multiple biopsies on the same patient. Forward stepwise model selection procedures were implemented. Variable selection criteria included $p < 0.15$, area under the receiver operating characteristic(ROC) curve(AUC) ≥0.005, or quasi-likelihood under the independence model criterion (QIC) with threshold of zero [5]. Final models were compared to identify variables that were robust to selection procedures. We first identified a full model including clinical predictors and 4Kpanel, and then a base model with serum PSA substituted for the 4Kpanel. In some clinics, prostate volume may not be reliably available, so models without prostate volume were fitted sequentially.

2.3.2. Model validation

Calibration plots were used to gauge the goodness of fit of each model. We used ROC analyses and AUC to assess the discriminatory capacity of a model for separating patients with and without reclassification. Decision curve analysis (DCA) was used to report the clinical net benefit of each model compared to biopsy-all and biopsy-none strategies [6]. The potential clinical impact was illustrated by plotting the number of cancers missed versus the number of biopsies avoided per 1000 individuals. To illustrate the clinical consequence of each model, we report the number of biopsies that could be avoided and the number of Gleason ≥7 cancers that might be missed if a risk-based threshold is applied as a criterion for biopsy. All evaluations were conducted on the initial biopsy and subsequent biopsy groups separately and combined. Confidence intervals (CIs) and significance tests were calculated using the bootstrap resampling procedure to account for within-subject correlations. All analyses were conducted using R version 3.1.1 (www.r-project.org).

3. Results

Of the 718 men in this study, there were 478 participants in the initial biopsy group for whom kallikreins were assayed: 319 in the training set (60 [18.8%] with Gleason ≥7) and 159 in the test set (34 [21.4%] with Gleason ≥7; Table 1). In bivariate analyses, prostate volume, ratio of positive to total cores, and the 4Kpanel were significantly associated with grade reclassification. There were 444 participants (of whom 204 were also in the initial biopsy group) with 633 subsequent surveillance biopsies, 422 in the training set (70 [17%] with Gleason ≥7; Table 2) and 211 in the test set (31 [15%] with Gleason ≥7; Supplementary Table 1). Biopsies in this group ranged from the second to eighth after diagnosis, and most patients had Gleason score 6 or no cancer at their surveillance biopsies, varying slightly across biopsy number.

In the full clinical model (Table 3) including the 4Kpanel, significant predictors for reclassification were BMI (odds ratio [OR] 1.09, 95% CI 1.04–1.14], >20% of cores positive in the prior biopsy (OR 2.10, 95% CI 1.33–3.32), a history of two or more biopsies negative for cancer (OR 0.19, 95% CI 0.04–0.85), prostate volume (per fold increase, OR 0.47, 95% CI 0.31–0.70), and 4Kpanel (OR 1.5, 95% CI 1.31–1.81). In the clinical model with serum PSA replacing the 4Kpanel, PSA was significantly associated with reclassification (per fold increase, OR 2.11, 95% CI 1.53–2.91) and age was not. In models that did not include prostate volume, the effects were similar for covariates left in the model (Supplementary Table 2). Model calibration in the test set showed predicted probabilities of reclassification closely matching the empirical rates (Supplementary Fig. 1).

Table 2 – Biopsy characteristics at each sequential surveillance biopsy after diagnosis for 558 participants in the training set

Parameter	Initial biopsy	Subsequent surveillance biopsies						
	First	Second	Third	Fourth	Fifth	Sixth	Seventh	Eighth
Biopsies (*n*)	319	246	108	34	20	10	3	1
CR for previous biopsy[a]								
Median (IQR)	0.08 (0.08)	0.07 (0.17)	0.08 (0.17)	0.06 (0.12)	0.06 (0.12)	0 (0.07)	0.11 (0.06)	0 (0)
Missing, *n* (%)	0	5 (2)	5 (5)	0	0	0	0	0
Median MCR[b] (IQR)	0.08 (0.08)	0.11 (0.08)	0.13 (0.15)	0.17 (0.13)	0.10 (0.17)	0.14 (0.15)	0.17 (0.08)	0.17 (0.00)
Negative biopsies[c], *n* (%)								
0	319 (100)	145 (59)	44 (41)	10 (29)	4 (20)	1 (10)	1 (33)	0
1	0	101 (41)	38 (35)	13 (38)	6 (30)	3 (30)	2 (67)	0
2	0	0	26 (24)	6 (18)	3 (15)	1 (10)	0	1 (100)
3	0	0	0	5 (15)	2 (10)	3 (30)	0	0
4	0	0	0	0	5 (25)	2 (20)	0	0
Median PV, cm^3 (IQR)	41.0 (26.5)	38.0 (27.0)	41.0 (27.0)	48.5 (25.0)	59.5 (36.5)	43.5 (27.8)	41.0 (19.5)	97.0 (0.0)
Biopsy GS, *n* (%)								
Negative	107 (34)	95 (39)	38 (35)	11 (32)	8 (40)	6 (60)	2 (67)	0
6	152 (48)	108 (44)	48 (45)	21 (62)	10 (50)	3 (30)	1 (33)	1 (100)
7	58 (18)	42 (17)	21 (19)	2 (6)	2 (10)	1 (10)	0	0
8	1 (0)	1 (0)	1 (1)	0	0	0	0	0
9	1 (0)	0	0	0	0	0	0	0

CR = core ratio; IQR = interquartile range; MCR = maximum CR; PV = prostate volume; GS = Gleason score.
[a] CR is defined as the number of biopsy cores containing cancer divided by the total number of biopsy cores in the previous biopsy.
[b] MCR among all previous biopsies.
[c] Number of surveillance biopsies in which no cancer was found.

Table 3 – Summary of fitted models including clinical variables + serum PSA or 4Kpanel in the training set

Variable	PSA + full clinical model		4K + full clinical model	
	OR (95% CI)	*p* value	OR (95% CI)	*p* value
Age	1.03 (1.00–1.06)	0.068		
Body mass index	1.11 (1.06–1.16)	<0.001	1.09 (1.04–1.14)	<0.001
Positive ore ratio >0.2	2.19 (1.39–3.44)	0.001	2.10 (1.33–3.32)	0.001
Negative biopsies ≥2	0.19 (0.04–0.80)	0.023	0.19 (0.04–0.85)	0.029
Log(prostate volume)	0.31 (0.20–0.48)	<0.001	0.47 (0.31–0.70)	<0.001
Log(PSA)	2.11 (1.53–2.91)	<0.001		
4Kpanel			1.54 (1.31–1.81)	<0.001

PSA = prostate-specific antigen; OR = odds ratio; CI = confidence interval.

ROC curve analysis (Table 4, Supplementary Fig. 2) comparing the full model with the 4Kpanel and the full clinical model with serum PSA indicated that the 4Kpanel significantly improved the accuracy for predicting reclassification (AUC 0.78 vs 0.74) in the initial surveillance biopsy, with a significant incremental value in AUC of 0.04 (95% CI 0.003–0.09). In a model without prostate volume, the incremental value in AUC was 0.07 (95% CI 0.02–0.11). The 4Kpanel did not improve prediction of reclassification in subsequent biopsies relative to PSA (AUC 0.75 vs 0.76).

Similar findings were observed in DCA. Compared to a clinical model with serum PSA, the model with 4Kpanel showed a higher net benefit for the initial surveillance biopsy, but there was no benefit for subsequent biopsies. All models showed substantial gain in net benefit compared with the biopsy-all and biopsy-none strategies across

Table 4 – Results of final regression models for reclassification

Base model	Area under the curve (95% confidence interval)		
	4K + clinical model	PSA + clinical model	Difference
Full clinical model			
Initial biopsy	0.783 (0.691–0.871)	0.740 (0.652–0.828)	0.043 (0.003–0.086)
Subsequent biopsy	0.754 (0.657–0.838)	0.755 (0.653–0.841)	−0.001 (−0.037–0.041)
Clinical model without prostate volume			
Initial biopsy	0.748 (0.654–0.840)	0.678 (0.579–0.774)	0.069 (0.016–0.114)
Subsequent biopsy	0.738 (0.633–0.825)	0.718 (0.611–0.810)	0.02 (−0.023–0.07)

PSA = prostate-specific antigen.
Confidence intervals were calculated with bootstrap accounting for correlations among individuals.

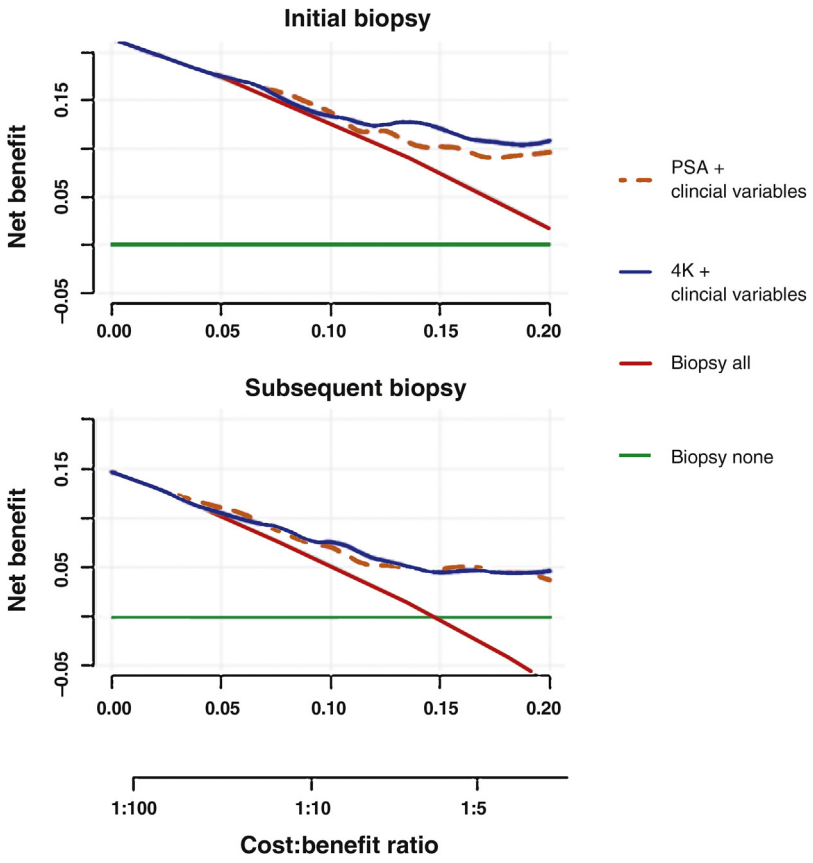


Fig. 1 – Decision curve analysis for full models with serum Prostate-specific antigen (PSA) or with the 4Kpanel. Strategies for biopsying all men (biopsy all) or no men (biopsy none) are also shown. The line with the highest net benefit at any particular threshold probability for biopsy (*x*-axis) will yield the best clinical results.

a range of plausible cost and benefit ratios (Fig. 1 and Supplementary Fig. 3).

The clinical consequences, or the number of biopsies and the number of high-grade cancers that could be avoided or delayed per 1000 patients, were illustrated based on prediction models with the 4Kpanel or PSA (Table 5). For example, using a model with the 4Kpanel and a clinical rule of only performing an initial surveillance biopsy in patients whose risk of high-grade cancer exceeded 10%, 252 biopsies would be avoided, 19 of which would contain high-grade cancer as defined by any pattern 4 disease, and zero biopsies with primary Gleason 4. Comparing the two models at the same numbers of biopsies avoided (Supplementary Fig. 4) shows that the 4K model appears to miss fewer higher-grade cancers while avoiding the same number of initial biopsies.

4. Discussion

In this study using a prospectively enrolled multi-institutional cohort of men on active surveillance, we show that addition of a panel of four kallikrein markers to a model that includes clinical information can significantly improve prediction of the outcome in the first surveillance biopsy. Both models performed comparably for prediction of reclassification in subsequent biopsies. Importantly, in DCA both models showed a higher net benefit compared to biopsy-all and biopsy-none strategies. Lastly, we showed that the 4Kpanel added to currently available clinical metrics and how the results impact clinical management.

There is a growing body of evidence that true Gleason 6 prostate cancer is indolent and will not cause harm if left untreated [10–12]. This knowledge is balanced by the known undersampling in prostate needle biopsies, and while some have advocated that select Gleason 3 + 4 cancers may undergo surveillance, level 1 clinical trial data and treatment guidelines generally recommend treatment of higher-grade cancers, including Gleason 3 + 4 disease [13,14]. Our efforts focus on developing tools for use after diagnosis of Gleason 6 prostate cancer to provide a higher degree of certainty that no occult high-grade cancer was missed at diagnosis. More accurate tools would not only support the practice of active surveillance but could also promote less intensive monitoring regimens.

A panel of four kallikreins, when combined in a mathematical algorithm, improves the prediction of newly diagnosed high-grade (Gleason $\geq$7) cancer [3]. This panel of markers also improved long-term prediction of metastatic disease among men with PSA $\geq$2 in a Swedish cohort [15]. In this study, we asked whether the same panel of markers [3] improved the prediction of high-grade disease in surveillance biopsies of men already diagnosed with Gleason

Table 5 – Clinical consequences showing the number of biopsies that could be avoided for initial surveillance biopsy or subsequent surveillance biopsy

HGC probability	Biopsies		High-grade cancers		Primary Gleason 4 cancers	
	Performed	Avoided	Found	Missed	Found	Missed
Initial surveillance biopsy						
Biopsy all	1000	0	214	0	44	0
Initial biopsy: risk by clinical variables + PSA						
>5%	943 (896–970)	57 (30–104)	214 (157–284)	0 (0–24)	44 (21–88)	0 (0–24)
>10%	761 (689–821)	239 (179–311)	201 (146–270)	13 (3–45)	44 (21–88)	0 (0–24)
>15%	509 (432–586)	491 (414–568)	164 (114–229)	50 (26–96)	38 (17–80)	6 (1–35)
Initial biopsy: risk by clinical variables + 4K						
>5%	956 (912–979)	44 (21–88)	214 (157–284)	0 (0–24)	44 (21–88)	0 (0–24)
>10%	748 (676–809)	252 (191–324)	195 (141–263)	19 (6–54)	44 (21–88)	0 (0–24)
>15%	522 (445–598)	478 (402–555)	182 (130–250)	31 (14–71)	44 (21–88)	0 (0–24)
Subsequent surveillance biopsies						
Biopsy all	1000	0	147	0	47	0
Risk by clinical variables + PSA						
>5%	844 (789–886)	156 (114–211)	147 (105–201)	0 (0–18)	47 (26–85)	0 (0–18)
>10%	692 (627–750)	308 (250–373)	133 (93–185)	14 (5–41)	43 (23–79)	5 (1–26)
>15%	445 (380–513)	555 (487–620)	109 (74–158)	38 (19–73)	43 (23–79)	5 (1–26)
Risk by clinical variables + 4K						
>5%	848 (794–890)	152 (110–206)	142 (101–196)	5 (1–26)	47 (26–85)	0 (0–18)
>10%	654 (588–715)	346 (285–412)	133 (93–185)	14 (5–41)	47 (26–85)	0 (0–18)
>15%	408 (344–475)	592 (525–656)	100 (66–147)	47 (26–85)	38 (19–73)	9 (3–34)

HGC = high-grade cancer.
Results are presented as the number (95% confidence interval) per 1000 men.

6 cancer. We found that when the kallikreins were assessed before the initial surveillance biopsy (sometimes called the confirmatory biopsy), the 4Kpanel provided incremental benefit for prediction of high-grade cancer (Gleason $\geq$7) over the clinical factors that are available at diagnosis. Specifically, depending on the choice from the various cutpoints that are based on the risk of high-grade disease, a substantial number of biopsies could be avoided while minimizing the number of missed high-grade cancers, few of which had primary pattern 4. The 4Kpanel was not of value over PSA for the prediction of reclassification in subsequent biopsies after the first surveillance biopsy. We found that the impact of other biopsy information, primarily volume of core involvement in previous biopsies and the number of previous negative biopsies, carries such a statistical weight in modeling that the impact of the 4Kpanel is minimized. For example, if a patient had low-volume disease at the initial surveillance biopsy or had subsequent negative biopsies after the initial diagnosis, then these factors were highly protective against biopsy reclassification at subsequent biopsy. It should be noted that our analysis of these subsequent biopsies used the 4Kpanel from the plasma sample that was closest to the subsequent biopsy, not necessarily the plasma sample from study entry, which could be months or years earlier than the subsequent biopsy.

We included serum PSA and prostate volume separately in our models instead of calculating PSA density, as we find a better model fit when the variables enter the model independently. While transurethral ultrasound prostate volume measurements may suffer from imprecision [16], statistical models that included prostate volume appeared to provide slightly improved predictive performance (AUC for all groups 0.77 with volume vs 0.75 without volume). Furthermore, prostate volume is a strong predictor of finding higher-grade cancers, with larger prostates being protective, as previously reported [17].

This study has limitations that merit mention. First, the model was developed and tested in the same cohort and with relatively limited numbers that resulted in wide confidence intervals and minor differences between the training and test sets. The results should clearly be validated in other cohort before clinical application. However, we expect that our results will be similar to those found in a community setting, as PASS is a multicenter center study that represents a broad spectrum of men utilizing active surveillance. Similarly, as PASS is primarily a Caucasian cohort, the findings of this study may not be generalizable to African American patients. Another limitation is that the serum PSA measurements used were obtained as part of standard clinical care, and the local site assays may differ from the one used with the 4Kpanel. Thus, the comparative modeling using PSA versus 4Kpanel may have slightly different tPSA values, with caution suggested for comparisons between the models. Lastly, as the use of imaging such as multiparametric MRI (mpMRI) is increasing, we do not have MRI data for most of our participants and recognize the potential value of future studies incorporating results from mpMRI and biomarkers in active surveillance.

5. Conclusions

The 4Kpanel was significantly associated with reclassification at the first surveillance biopsy, providing incremental

value over routine clinical information, and the 4K model performed significantly better than the base model in this group. The 4Kpanel did not add predictive value to a PSA clinical model for biopsy decision-making for men at subsequent surveillance biopsies. This work aims to provide clinical validation of a biomarker that will help determine those men who have or will develop aggressive prostate cancer, allowing for the accurate determination of those men who may avoid or delay the burden of immediate treatment safely, while concurrently identifying men who may optimally benefit from early treatment.

Author contributions: Daniel W. Lin had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Lin, Newcomb, Sjoberg, Dong, Zheng.
Acquisition of data: Lin, Newcomb, Brooks, Carroll, Cooperberg, Dash, Ellis, Fabrizio, Gleave, Morgan, Nelson, Thompson, Zheng.
Analysis and interpretation of data: Lin, Newcomb, Brown, Sjoberg, Dong, Nelson, Thompson, Zheng.
Drafting of the manuscript: Lin, Newcomb, Brown, Sjoberg, Dong, Nelson, Thompson, Zheng.
Critical revision of the manuscript for important intellectual content: Lin, Newcomb, Brown, Sjoberg, Dong, Brooks, Carroll, Cooperberg, Dash, Ellis, Fabrizio, Gleave, Morgan, Nelson, Thompson, Zheng.
Statistical analysis: Brown, Sjoberg, Dong, Zheng.
Obtaining funding: Lin, Newcomb, Nelson, Zheng.
Administrative, technical, or material support: Lin, Nelson.
Supervision: Newcomb, Lin, Nelson.
Other: None.

Financial disclosures: Daniel W. Lin certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Yan Dong is an employee of OPKO. The remaining authors have nothing to disclose.

Funding/Support and role of the sponsor: This work was supported by the Department of Defense (PC130355), the Canary Foundation, and the Institute for Prostate Cancer Research. The sponsors played a role in the design and conduct of the study, and collection, management, and analysis of the data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.eururo.2016.11.017.

References

[1] Ankerst DP, Xia J, Thompson Jr IM, et al. Precision medicine in active surveillance for prostate cancer: development of the Canary-Early Detection Research Network active surveillance biopsy risk calculator. Eur Urol 2015;68:1083–8.
[2] Parekh DJ, Punnen S, Sjoberg DD, et al. A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. Eur Urol 2015;68:464–70.
[3] Bryant RJ, Sjoberg DD, Vickers AJ, et al. Predicting high-grade cancer at ten-core prostate biopsy using four kallikrein markers measured in blood in the ProtecT study. J Natl Cancer Inst 2015;107(7).
[4] Newcomb LF, Thompson Jr IM, Boyer HD, et al. Outcomes of active surveillance for clinically localized prostate cancer in the prospective, multi-institutional Canary PASS cohort. J Urol 2016;195:313–20.
[5] Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;57:120–5.
[6] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.
[7] US Food and Drug Administration. Innovation or stagnation: challenge and opportunity on the critical path to new medical products. Washington, DC: US Department of Health and Human Services; 2004.
[8] McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 2005;97:1180–4.
[9] Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst 1996;88:1456–66.
[10] Ross HM, Kryvenko ON, Cowan JE, Simko JP, Wheeler TM, Epstein JI. Do adenocarcinomas of the prostate with Gleason score (GS) $\leq$6 have the potential to metastasize to lymph nodes? Am J Surg Pathol 2012;36:1346–52.
[11] Eggener SE, Badani K, Barocas DA, et al. Gleason 6 prostate cancer: translating biology into population health. J Urol 2015;194:626–34.
[12] Carter HB, Partin AW, Walsh PC, et al. Gleason score 6 adenocarcinoma: should it be labeled as cancer? J Clin Oncol 2012;30:4294–6.
[13] Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. N Engl J Med 2012;367:203–13.
[14] Mohler JL, Kantoff PW, Armstrong AJ, et al. Prostate cancer, version 1.2014. J Natl Compr Canc Netw 2013;11:1471–9.
[15] Stattin P, Vickers AJ, Sjoberg DD, et al. Improving the specificity of screening for lethal prostate cancer using prostate-specific antigen and a panel of kallikrein markers: a nested case-control study. Eur Urol 2015;68:207–13.
[16] Rodriguez Jr E, Skarecky D, Narula N, Ahlering TE. Prostate volume estimation using the ellipsoid formula consistently underestimates actual gland size. J Urol 2008;179:501–3.
[17] Roobol MJ, van Vugt HA, Loeb S, et al. Prediction of prostate cancer risk: the role of prostate volume and digital rectal examination in the ERSPC risk calculators. Eur Urol 2012;61:577–83.

Refined analysis of prostate specific antigen kinetics to predict prostate cancer active surveillance outcomes

Matthew R. Cooperberg, MD, MPH[1,2]; James D. Brooks, MD[3]; Anna V. Faino, MS[4]; Lisa F. Newcomb, PhD[5,6]; James T. Kearns, MD[6]; Peter R. Carroll, MD, MPH[1]; Atreya Dash, MD[6]; Ruth Etzioni, PhD[4]; Michael D. Fabrizio, MD[7]; Martin E. Gleave, MD[8]; Todd M. Morgan, MD[9]; Peter S. Nelson, MD[10]; Ian M. Thompson, MD[11]; Andrew A. Wagner, MD[12]; Daniel W. Lin, MD[5,6]; and Yingye Zheng, PhD[4]

Author affiliations:
[1] Department of Urology, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco CA
[2]Department of Epidemiology & Biostatistics, University of California, San Francisco CA.
[3] Department of Urology, Stanford University, Stanford CA
[4] Biostatistics Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA
[5] Cancer Prevention Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA
[6] Department of Urology, University of Washington, Seattle WA
[7] Department of Urology, Eastern Virginia Medical School, Virginia Beach VA
[8] Department of Urologic Sciences, University of British Columbia, Vancouver BC
[9] Department of Urology, University of Michigan, Ann Arbor MI
[10] Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle WA
[11] CHRISTUS Medical Center Hospital, San Antonio TX
[12] Division of Urology, Beth Israel Deaconess Medical Center, Boston MA

Running title: PSA kinetics in prostate cancer active surveillance

Keywords: PSA, kinetics, prostate cancer, active surveillance, outcomes

Financial Support: Department of Defense Prostate Cancer Research Program (Grants W81XWH1110489 and W81XWH1410595); Canary Foundation

Corresponding author:
Matthew R. Cooperberg
UCSF Helen Diller Family Comprehensive Cancer Center
Box 1695, 550 16th Street
San Francisco, CA 94143-1695, USA
Tel. +1 415 8853660
Fax: +1 415 8857443
E-mail address: mcooperberg@urology.ucsf.edu

The authors declare no potential conflicts of interest.

Word count: 2698
Figures and tables: 5 total (2 tables, 3 figures)

Abstract

Background:
For men on active surveillance for prostate cancer, utility of prostate specific antigen (PSA) kinetics in predicting pathologic reclassification remains controversial.

Objective:
We aimed to develop prediction methods for utilizing serial PSA and evaluate frequency of collection.

Design, Setting, and Participants:
Data were collected from men enrolled in the multicenter Canary Prostate Active Surveillance Study(PASS), among whom PSAs and biopsies were performed on pre-specified schedules. We developed a PSA kinetic parameter (PSAk) based on a linear mixed effect model (LMEM) that accounted for serial PSA levels.

Outcome Measurements and Statistical Analysis:
The association of diagnostic PSA and/or PSAk with time to reclassification (increase in cancer grade and/or volume) was evaluated using multivariable Cox proportional hazards models.

Results and Limitations:
851 men met study criteria; 255 (30%) had a reclassification event within 5 years. Median follow-up was 3.0 years. After adjusting for prostate size, time since diagnosis, biopsy parameters, and diagnostic PSA, PSAk was a significant predictor of reclassification (HR for each 0.10 increase in PSAk = 1.6 (95% CI 1.2-2.1, p< .001). The PSAk model improved stratification of risk prediction for the top and bottom deciles of risk over a model without PSAk. Model performance was essentially identical using PSAs measured every 6 months rather than 3 months. The major limitation is the reliability of reclassification as an endpoint, although it drives most treatment decisions.

Conclusions:
PSAk calculated using LMEM statistically significantly predicts biopsy reclassification. Models that use repeat PSA measurements outperform a model incorporating only diagnostic PSA. Model performance is similar using PSA assessed every 3 or 6 months. These results will inform optimal incorporation of PSA trends into active surveillance protocols and risk calculators.

Patient Summary:
In this report we looked at whether repeat PSA measurements, or PSA kinetics, improve prediction of biopsy outcomes in men using active surveillance to manage localized prostate cancer. We found that in a large multi-center active surveillance cohort, PSA kinetics improves the prediction of surveillance biopsy outcome.

Introduction

Given the prolonged natural history and indolent behavior of most low-risk prostate cancers, active surveillance (AS) has been developed as an alternative to immediate treatment. Surveillance is now recognized as a preferred strategy for low-risk disease(1) and is offered to a large and growing proportion of men, both in the U.S.(2,3) and internationally.(4) While substantial variation persists in terms of eligibility criteria for surveillance, followup intervals, and triggers for intervention, all AS protocols are based principally on repeated prostate specific antigen (PSA) measurements and periodic re-biopsy.(1)

How to collect and interpret serial PSA data optimally in the AS setting, however, remains unclear. In most centers, PSA is collected quarterly, with the goal of identifying men with a rapid PSA rise that may signify aggressive disease. However, studies to date have not shown analyses of PSA kinetics to be informative in most cases. In multiple cohorts, PSA kinetics consistently failed to predict reclassification based on biopsy parameters (i.e., increase in biopsy Gleason grade and/or tumor volume).(5,6) In the prospective, multicenter Canary Prostate Active Surveillance Study (PASS), PSA doubling time <36 months was originally a criterion for progression, but since consistently few men met this threshold it was dropped from the protocol.(7)

A limiting factor in most AS cohorts reporting outcomes is the relatively short duration of followup and limited longitudinal PSA data. As the PASS cohort has matured with longer followup, additional PSA measurements, and more reclassification events, we have an

opportunity to determine the extent to which PSA kinetics might facilitate improved decision-making for men on surveillance for low-risk prostate cancer. We also aimed to determine whether quarterly PSA measurements are necessary for accurate assessment of PSA kinetics, or whether semiannual measurement would be sufficient.

Methods

The Canary Prostate Active Surveillance Study (PASS) is a multicenter, prospective cohort enrolling men on AS at nine North American centers. Men eligible for AS provide informed consent under institutional review board supervision (clinicaltrials.gov NCT00756665). In PASS, PSA is measured every 3 months, clinic visits occur every 6 months, and ultrasound-guided biopsies are performed 6 to 12 months after diagnosis, 24 months after diagnosis, then every 2 years. Other tests, including magnetic resonance imaging, are performed at the clinicians' discretion, but as enrollment started in 2008, the majority of men did not undergo these procedures. For the current study, participants were enrolled before February 2016, had diagnostic Gleason grade ≤3+4 and <34% of biopsy cores involved with cancer, no history of 5α-reductase inhibitor use, and at least one PSA and one biopsy following diagnosis. The primary outcome was tumor reclassification, defined as an increase in primary or secondary Gleason grade, or increase in tumor volume to ≥34% of total biopsy cores involved.

Statistical analysis

PSA may be measured irregularly during AS, and is characterized by within-individual random variation, which may attenuate associations between PSA kinetics and clinical outcomes. To study longitudinal PSA measurements as predictors of reclassification while accommodating

these complicating factors, a two-stage procedure was used.(8,9) Through this process we derived a novel PSA kinetic (PSAk) parameter, which we treated like a biomarker, and our approach conformed to the REMARK criteria for novel biomarkers.(10)

First, we calculated PSAk using a linear mixed effect model (LMEM), in which the natural logarithm of PSA (ln(PSA)) was modeled as a linear function of time since diagnosis, with a random intercept indicating the individual-specific ln(PSA) at diagnosis, and a random slope reflecting individual-specific rate of change over time. PSAk for each participant based on all his PSA measurements from diagnosis to a specific observation time was derived using the best linear unbiased predictor (BLUP) estimator from the LMEM (see Supplemental Methods). Intraclass correlation (ICC) was calculated to assess how much of the variability in PSA was explained by between-participant variance compared to total variance. A high ICC indicates strong correlations among PSA measurements from the same individual.

Two other approaches for calculating PSA kinetics were considered: a linear regression model using all PSA measurements from diagnosis to an observation time (simple PSAk, or PSAkS), and a slope change using 2 PSA measurements closest to and including the observation time (restricted simple PSAk, or PSAkRS). Models were adjusted for prostate size.

Second, Cox proportional hazards (PH) models were used to determine risk of future reclassification as a function of covariates at each observation time. The outcome was defined as time from each PSA measurement to reclassification or censoring. Participants were

censored at treatment, last study contact, or four years after biopsy. Individual-specific PSAk at each measurement time estimated from stage one was the key covariate. Other covariates considered were: age, ln(prostate size), ln(observation time since diagnosis), diagnostic Gleason (3+3 or 3+4), percent of biopsy cores positive, number of biopsies since diagnosis (0,1,2,3, or 4+), negative biopsy since diagnosis, recent biopsy result (cancer versus no cancer), and ln(diagnostic PSA).

Hazard ratios (HR) and 95% Confidence Intervals (CI) were calculated with robust variance estimates to account for correlations from multiple observations from the same individual. Model fit was compared using the Akaike information criterion (AIC); smaller AIC indicated better goodness of fit. Non-significant variables were backwards eliminated using a p-value cutoff of 0.05. A sensitivity analysis was performed including only participants whose biopsies were all per-protocol (on-time).

To assess performance of the multivariable model incorporating PSAk, a Cox PH model was used to predict subsequent 3 year outcomes from 1 year after diagnosis for each participant, using PSA data from diagnosis to 1 year. Time-dependent receiver operating characteristic (ROC) curves and areas under the curve (AUC) were used to quantify the performance of models. Bootstrapping methods were used to obtain 95% confidence intervals for AUCs. To evaluate the usefulness of PSAk for risk stratification, we categorized the model-based risk into: lowest 10%, middle, and highest 10% risk. We compared reclassification-free probabilities

among these risk groups using a Kaplan-Meier (KM) analysis. Risk groups were generated using models with and without PSAk.

We ran an additional analysis comparing modeling performance for PSA measured every 6 versus every 3 months since diagnosis, and assessed whether semiannual and quarterly PSA measurement yielded similar predictions. A two-sided p-value <0.05 was considered significant for analyses, which were performed using R version 3.3.0.

Results

There were 851 men in PASS that were included. Median (interquartile range [IQR]) follow up was 3.0 (1.7-5.1) years. Among all participants, 291 (34%) were reclassified by an increase in biopsy Gleason grade or tumor volume to ≥34% of total biopsy cores with cancer, of which 210 (25%) and 255 (30%) reclassified within 3 and 5 years of diagnosis, respectively. The median participant age was 62. Six percent were African-American and 4% other non-Caucasian race. Eighty percent of biopsies were per-protocol (on-time), 11% early, and 9% late. Table 1 summarizes the clinical characteristics of the cohort. Compared to censored participants, reclassified participants had similar clinical risk as assessed by the CAPRA score (p = 0.95), but had smaller prostates, higher PSAD, and a higher proportion of diagnostic biopsy cores involved (all p<0.001).

The annual percent change in PSA estimated by the LMEM was 4.3 (95% CI 3.4-5.2, p<0.001). As determined by the ICC, 85% of the observed variation in PSA kinetics was explained by

between-participant variation, and 15% by within-participant variation. By LMEM estimation, PSA rose 8.1% annually (95% CI 6.0-10.3, p<0.001) for reclassified participants, and 0.8% (95% CI -0.7-2.4, p=0.33) for censored participants (Figure 1).

In a Cox PH model including both ln(diagnostic PSA) and PSAk, PSAk was independently associated with reclassification, with HR 1.4 (95% CI 1.1-1.9) for each 0.1 unit increase. HR for ln(diagnostic PSA) at diagnosis was 1.3 (95% CI 1.0-1.6). In a multivariable model adjusted for prostate size, time since diagnosis, percent of biopsy cores involved in the most recent biopsy, and any negative biopsy after diagnosis, the HRs for ln(diagnostic PSA) and PSAk remained significant: 1.6 (95% CI 1.3-2.1) and 1.6 (95% CI 1.2-2.1), respectively.

In a secondary analysis modeling PSA kinetics calculated using three different methods, a restricted simple PSAk (see methods) was not associated with time to reclassification (p = 0.11). The association between simple PSAk from linear regression and time to reclassification was less significant (p = 0.004, compared to p = 0.001 for PSAk) and not as strong (HR = 1.08, 95% CI 1.02-1.14 for each 0.1 unit increase) as PSAk based on the best linear unbiased predictor (BLUP) estimator from the LMEM (HR=1.53, 95% CI 1.19-1.98 for each 0.1 unit increase). The model with PSAk had the highest goodness-of-fit with respect to AIC. In a model that contained all three PSA kinetic measurements, PSAk was statistically significant while PSAkS and PSAkRS were not (Table 2). Thus, the simple methods of calculating PSA kinetics were not considered further. No meaningful differences in parameter estimates or statistical significance were observed in the sensitivity analysis including only participants with per-protocol biopsies.

The AUC for the full multivariable model including PSAk in predicting 3 year reclassification outcomes from a measurement time of 1 year after diagnosis was 0.79 (95% CI 0.74-0.84). As illustrated in Figure 2, when subgroups of low-, middle- and high-risk for reclassification were identified based on models with and without PSAk, the inclusion of PSAk was better able to distinguish extreme subgroups of individuals (10% of the cohort each) with low and high event rates in the years after the prediction. The reclassification-free probability based on the KM estimator in the low risk group at 4 years after the 1-year measurement time was 1.00 (95% CI 1.00-1.00) with PSAk and 0.95 (95% CI 0.88-1.00) without PSAK in the model. In contrast, the reclassification-free probability in the high risk group at 4 years after the 1-year measurement time was 0.21 (95% CI 0.10-0.43) with PSAk versus 0.34 (95% CI 0.21-0.54) without PSAk.

Calculating PSAk based on semiannual rather than quarterly PSA measurements yielded tightly correlated results: $r=0.95$, $p<0.001$ (Figure 3). Recalculating the multivariable Cox PH model described above using only semiannual PSAs yielded substantially similar results, with the new HRs for ln(PSA) and PSAk 1.6 (95% CI 1.2-2.0) and 1.8 (95% CI 1.3-2.5), respectively. The AUCs for 3 and 6 month models are similar (Supplemental Figure 1).

Discussion

PSA kinetics have long been studied as indicators of prostate cancer prognosis, at decision points ranging from whether a man should undergo initial prostate biopsy(11) to early identification of advanced disease progression.(12) The utility of PSA kinetics in the pre-

treatment setting has been difficult to establish for a number of reasons, including close correlation with static PSA at diagnosis,(13) relatively short followup and limited longitudinal data in most series, PSA "noise" from non-cancer sources, and the myriad published definitions of PSA kinetics, such as velocity, doubling time, and other measures of growth.(14) We found stronger strength of association and better prediction calibration when PSA kinetics was calculated based on a LMEM that accounted for both the general trend of increasing PSA over time in the cohort and individual-specific trajectories, while discounting the random noise in the PSA measurements.

For men on AS for low-risk prostate cancer, a rapidly rising PSA intuitively would seem to predict aggressive disease and adverse outcomes. In fact, in the Toronto cohort, one of the earliest AS cohorts in North America, a PSA doubling time (DT) <2 years was initially the primary trigger for intervention. This threshold was found inadequately sensitive and was extended to <3 years. However, while the definition of rapid PSADT in this cohort does predict outcomes after definitive treatment for men initially on AS, it was found to be non-specific and is now considered in the context of other indicators, particularly grade reclassification.(15) In two other large, single-institution cohorts in the US, PSA kinetics have not been proven useful with relatively short-term followup. In the UCSF cohort, PSADT <3 years was associated with increased risk of reclassification—but only one man in the first 241 enrolled met this threshold.(6) In the Johns Hopkins cohort, PSADT and PSA velocity, calculated as PSA multiplied by the slope of a linear regression of log(PSA), were both poor predictors of reclassification, with AUCs of 0.59 and 0.61, respectively.(5) In this cohort, however, enrollment criteria for AS

are very restrictive, yielding a narrow dynamic range in terms of progression risk in which to evaluate PSA kinetics.

In this study, we analyzed PSA kinetics in a multicenter cohort with PSA data collected at protocol-mandated intervals, relatively long follow up, and centralized analysis. We employed an analytic strategy which allowed the models to account for prior PSA history at each individual PSA measurement in an individual participant's trajectory while borrowing information from the general trend across all participants and accommodating for random variability in PSA. In a plot of individual PSA trajectories, a higher slope overall was found for those who did versus did not reclassify (Figure 1). Moreover, the addition of PSAk to a rich multivariable model improved performance of the model, suggesting that PSAk may be considered as an additional biomarker for outcomes on AS. In general, this finding suggests that collecting PSA measurements over time to provide an updated outcome is clinically useful, and our approach of calculating PSAk provides a summary that effectively reflects the changes of PSA over time. On the other hand, given essentially identical results analyzing every 3 month vs. every 6 month PSA data, we can conclude that in most cases PSA does not need to be measured any more often than semiannually.

The imaging and molecular tests available to supplement standard clinical data to guide decision making for men with low-risk disease are proliferating rapidly, and the potential clinical utility of PSAk should be considered in this context. We have adhered to REMARK criteria for biomarker reporting(10) to as great an extent as possible. In particular, we stress

that all PSA and outcome data have been collected and reported prospectively throughout the duration of PASS, and all analyses conducted centrally. Although several biomarkers, as well as multiparametric MRI, are currently marketed for decision-making with respect to AS,(16) none so far has been validated in a prospective AS cohort. Moreover, PSAk has the advantage of requiring no additional biomaterial, nor any incremental cost.

A few caveats should be noted. PSA levels are reported directly from the Canary/PASS clinical sites, and reflect different laboratories. We therefore cannot control for inter-assay variability in PSA levels. However, men are instructed to use the same lab consistently for their PSA measurements, and we expect assay variability would introduce bias toward a null rather than a false positive result. We examined performance of PSAk above and below the PSAD threshold of 0.15 to better understand the performance of PSAk relative to PSAD. Our finding of differential performance at high and low PSAD is intriguing—perhaps reflecting better information from PSA trends in the absence of substantial BPH—and merits further examination. Because changes in PSA may affect decisions regarding biopsy performance (and therefore opportunity to identify reclassification), a risk of ascertainment bias exists. However, compliance with biopsy schedules in PASS is generally excellent (80% of biopsies on-time), and the sensitivity analysis excluding men with non-compliant biopsies did not change the results.

Perhaps the most important limitation is the reliability of our endpoint. The principal question was the ability of PSA kinetics to predict biopsy reclassification. We acknowledge that reclassification itself is an imperfect endpoint, as it may reflect initial undersampling,(17)

variation in different pathologist interpretations,(18) and/or minimal changes in the tumor which have little clinical importance. However, our reclassification definition is consistent with those used by most other AS cohorts, and these changes do frequently drive treatment decision-making in contemporary practice. Therefore, while perhaps not biologically optimal, we believe our findings are quite relevant for current clinical management, and can in fact improve AS care.

In conclusion, we found that a sophisticated mathematical approach to measuring PSA kinetics, as reflected in the novel PSAk parameter, can improve prediction of outcomes for men on surveillance for prostate cancer, and that PSA needs to be measured no more often than semiannually. PSA should of course never be interpreted in a vacuum, and we did not identify a PSAk threshold which should always indicate treatment. However, these results suggest that PSAk or similar assessments of kinetics should be considered in future multivariable models of active surveillance outcomes.

References

1. Chen RC, Rumble RB, Loblaw DA, Finelli A, Ehdaie B, Cooperberg MR, et al. Active Surveillance for the Management of Localized Prostate Cancer (Cancer Care Ontario Guideline): American Society of Clinical Oncology Clinical Practice Guideline Endorsement. J Clin Oncol. 2016.

2. Womble PR, Montie JE, Ye Z, Linsell SM, Lane BR, Miller DC, et al. Contemporary use of initial active surveillance among men in Michigan with low-risk prostate cancer. Eur Urol. 2015;67:44–50.

3. Cooperberg MR, Carroll PR. Trends in Management for Patients With Localized Prostate Cancer, 1990-2013. JAMA. 2015;314:80–2.

4. Loeb S, Folkvaljon Y, Curnyn C, Robinson D, Bratt O, Stattin P. Uptake of Active Surveillance for Very-Low-Risk Prostate Cancer in Sweden. JAMA Oncol. 2016.

5. Ross AE, Loeb S, Landis P, Partin AW, Epstein JI, Kettermann A, et al. Prostate-specific antigen kinetics during follow-up are an unreliable trigger for intervention in a prostate cancer surveillance program. J Clin Oncol. 2010;28:2810–6.

6. Whitson JM, Porten SP, Hilton JF, Cowan JE, Perez N, Cooperberg MR, et al. The relationship between prostate specific antigen change and biopsy progression in patients on active surveillance for prostate cancer. J Urol. 2011;185:1656–60.

7. Newcomb LF, Thompson IM, Boyer HD, Brooks JD, Carroll PR, Cooperberg MR, et al. Outcomes of Active Surveillance for Clinically Localized Prostate Cancer in the Prospective, Multi-Institutional Canary PASS Cohort. J Urol. 2016;195:313–20.

8. Zheng Y, Heagerty PJ. Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. Biostatistics. 2004;5:615–32.

9. Maziarz M, Heagerty P, Cai T, Zheng Y. On longitudinal prediction with time-to-event outcome: Comparison of modeling options. Biometrics. 2016.

10. Mcshane L, Altman D, Sauerbrei W, Taube S, Gion M, Clark G. Reporting Recommendations for Tumor Marker Prognostic Studies. J Clin Oncol. 2005;23:9067.

11. Carter HB, Ferrucci L, Kettermann A, Landis P, Wright EJ, Epstein JI, et al. Detection of life-threatening prostate cancer with prostate-specific antigen velocity during a window of curability. J Natl Cancer Inst. 2006;98:1521–7.

12. Scher HI, Halabi S, Tannock I, Morris M, Sternberg CN, Carducci MA, et al. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. J Clin Oncol. 2008. pages 1148–59.

13. Vickers AJ, Till C, Tangen CM, Lilja H, Thompson IM. An Empirical Evaluation of Guidelines on Prostate-specific Antigen Velocity in Prostate Cancer Detection. J Natl Cancer Inst. 2011;103:462–9.

14. den Bergh van RCN, Roemeling S, Roobol MJ, Wolters T, Schröder FH, Bangma CH. Prostate-specific antigen kinetics in clinical decision-making during active surveillance for early prostate cancer--a review. Eur Urol. 2008;54:505–16.

15. Klotz L, Vesprini D, Sethukavalan P, Jethava V, Zhang L, Jain S, et al. Long-Term Follow-Up of a Large Active Surveillance Cohort of Patients With Prostate Cancer. J Clin Oncol. 2015;33:272–6.

16. Leapman MS, Nguyen HG, Cooperberg MR. Clinical Utility of Biomarkers in Localized Prostate Cancer. Curr Oncol Rep. 2016;18:30.

17. Porten SP, Whitson JM, Cowan JE, Cooperberg MR, Shinohara K, Perez N, et al. Changes in prostate cancer grade on serial biopsy in men undergoing active surveillance. J Clin Oncol. 2011;29:2795–800.

18. McKenney JK, Simko J, Bonham M, True LD, Troyer D, Hawley S, et al. The potential impact of reproducibility of Gleason grading in men with early stage prostate cancer managed by active surveillance: a multi-institutional study. J Urol. 2011;186:465–9.

Table 1: Participant characteristics

	All Participants n=851	**Reclassified Participants n=291**	**Censored Participants n=560**	**P-value***
Time to event/censor (years), median [IQR]	3.0 [1.7-5.1]	2.0 [1.1-3.2]	3.9 [2.4-5.4]	<.001
No. PSA values, median [IQR]	8 [4-13]	5 [3-9]	9 [5-15]	<.001
Dx PSA, median [IQR]	4.8 [3.6-6.3]	4.9 [3.9-6.3]	4.7 [3.5-6.3]	0.07
Prostate size, median [IQR]	40 [30-54]	35 [27-46]	44 [32-58]	<.001
Dx PSA density, median [IQR]	0.11 [0.08-0.16]	0.14 [0.10-0.18]	0.10 [0.07-0.14]	<.001
Dx cores percentage, median [IQR]~	8 [8, 17]	17 [8, 17]	8 [8, 17]	<.001
Dx age, median [IQR]	62 [57-67]	63 [58-67]	62 [57-67]	0.22
Dx CAPRA score, n(%)~				
0	30 (4)	9 (3)	21 (4)	
1	510 (64)	177 (64)	333 (64)	0.95
2	206 (26)	73 (26)	133 (25)	
3+	56 (7)	19 (7)	37 (7)	

PSA = prostate specific antigen, IQR = interquartile range, CAPRA = Cancer of the Prostate Risk Assessment, Dx = at diagnosis

* P-value from 2 sample t-test for continuous variables, and from chi-squared test for categorical variables. For time to event/censor, number of PSA values, PSA density and cores percentage, p-value from Wilcoxon test.

~ 28 participants and 49 participants are missing cores percentage and CAPRA score at diagnosis, respectively.

Table 2. Comparing simple PSAk (PSAkS), restricted simple PSAk (PSAkRS), and PSAk in Cox PH models, n=841~.

Variable	PSAkS model		PSAkRS model		PSAk model		Model containing all PSA kinetic measurements	
	HR [95% CI]	P-value	HR [95% CI]	P-value	HR [95% CI]	P-value	HR [95% CI]	P-value
Dx PSA	1.75 (1.34, 2.29)	<.0001	1.73 (1.33, 2.25)	<.0001	1.77 (1.34, 2.34)	<.0001	1.79 (1.35, 2.37)	<0.001
PSAkS (0.10 increase)	1.08 (1.02, 1.14)	0.004					1.00 (1.00,1.00)	0.208
PSAkRS (0.10 increase)			1.01 (1.00, 1.02)	0.11			1.00 (1.00,1.00)	0.247
PSAk (0.10 increase)					1.53 (1.19, 1.98)	0.001	1.51 (1.17, 1.95)	0.001
AIC	26587		26599		26492		26492	

~ Participants were required to have non-missing PSAkS, PSAkRS and PSAk to be considered in the model comparison. All models adjusted for prostate size.

Figure legends

Figure 1: PSA trajectory prior to reclassification or censoring.

In this spaghetti plot, individual ln(PSA) trajectories are plotted in red for reclassified participants and in blue for censored participants, omitting PSA data within 2 years prior of censor date to look at long-term non-events. Smoothed trend lines were added using LOESS. A separate LMEM analysis found a slope of 8.1%/year for reclassified participants and 0.8% for censored participants. (P-value for interaction between reclassification group and PSA change: <0.001).

Figure 2: Predicting reclassification outcomes

Kaplan-Meier plots showing reclassification-free probabilities at 4 years after the 1-year measurement time, using data up to 1 year after diagnosis. Left panel: model-based risk categories from Cox PH model adjusted for PSA at diagnosis, prostate size, time since diagnosis, most recent percent of biopsy cores involved, and history of any negative biopsy—but not PSAk. Right panel: similar analysis adjusted for the same variables in addition to PSAk. The PSAk model improved stratification of risk prediction for the top and bottom deciles of risk over a model without PSAk.

Figure 3: Quarterly vs. semiannual PSA measurement

Correlation between PSAk calculations based on every 3-month vs. every 6-month PSA measurement is illustrated. Pearson correlation (r) 0.95, p<0.001.

Figure 1

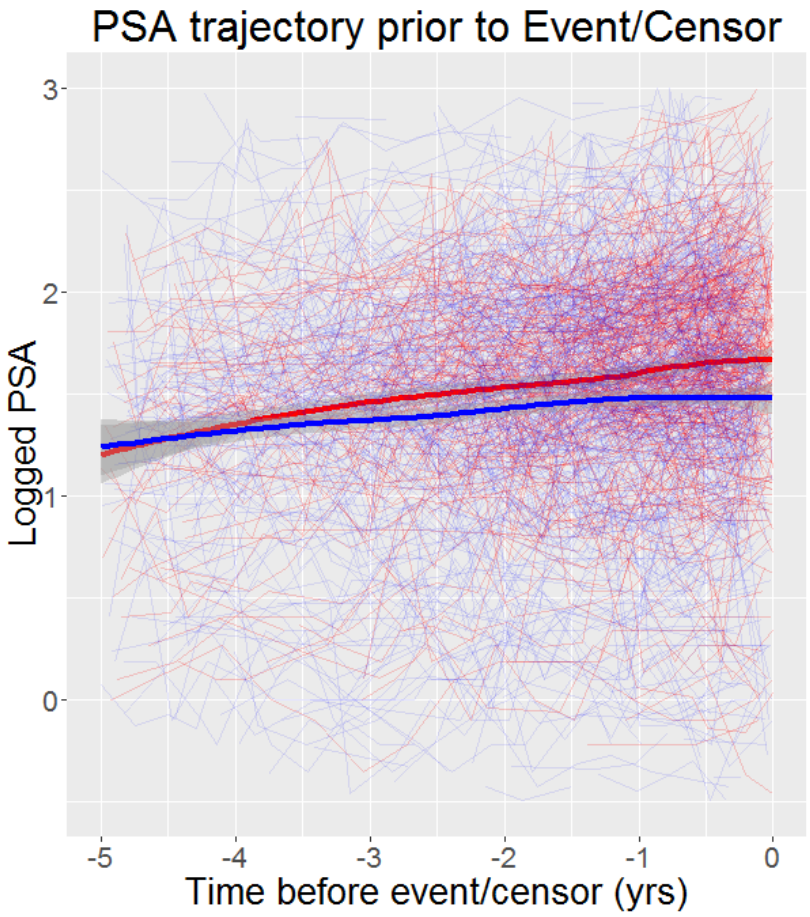
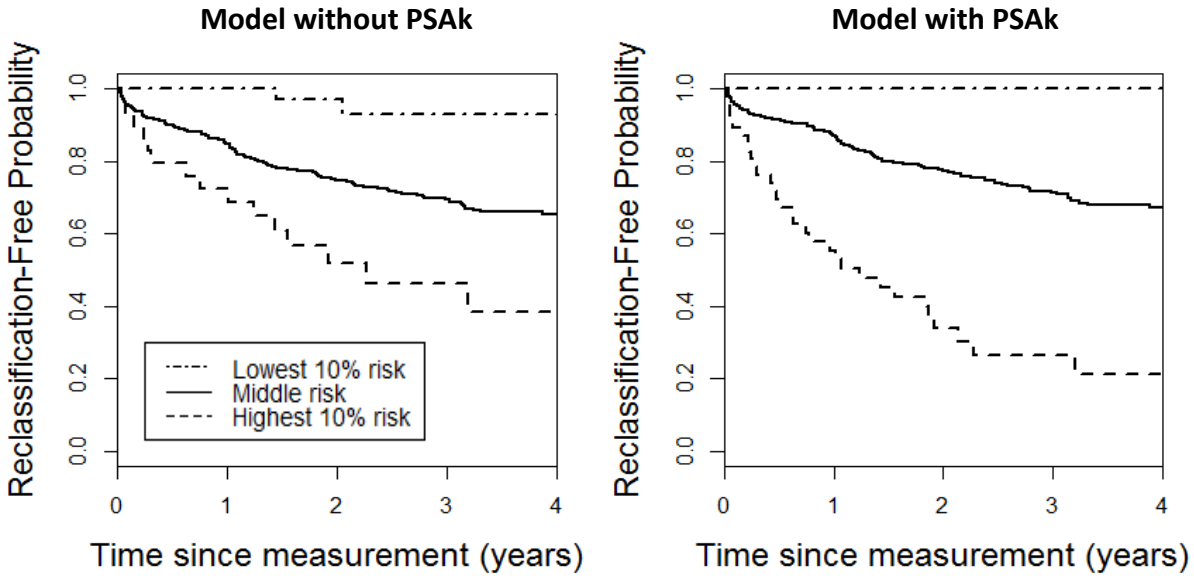

PSA trajectory prior to Event/Censor
Logged PSA
3
2
1
0
-5
-4
-3
-2
-1
0
Time before event/censor (yrs)

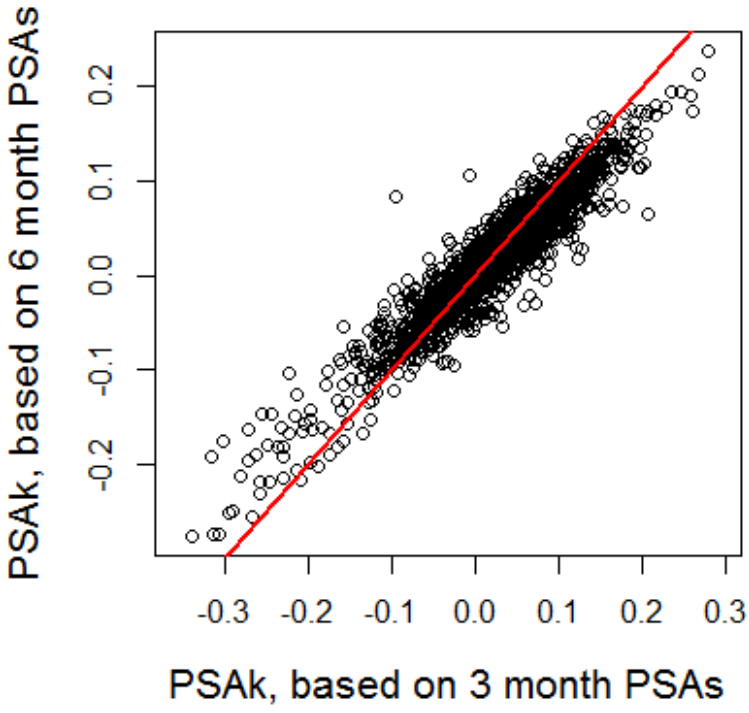
Figure 2



Model-Based Risk Group*	Reclassification-Free Probability at 4 years after the 1-year measurement time (95% CI)	
	Model without PSAk	Model with PSAk
Lowest 10% risk	0.95 (0.88, 1.00)	1.00 (1.00, 1.00)
Middle risk	0.66 (0.61, 0.72)	0.67 (0.62, 0.73)
Highest 10% risk	0.34 (0.21, 0.54)	0.21 (0.10, 0.43)

* Model-based risk is calculated at a measurement time of 1 year after diagnosis using all data available up to the measurement time.

Figure 3


PSAk, based on 6 month PSAs
-0.2
-0.1
0.0
0.1
0.2
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
PSAk, based on 3 month PSAs

Supplemental methods

Calculation of PSAk using BLUP

A participant's BLUP is time specific, and includes all PSA information up to and including the time of interest. The PSAk is calculated as follows for participant *i* at time *j*:

$$PSAkBLUP_{ij} = \beta_{t,fixed} + [0\ 1] * \hat{\beta}_s$$

where $\beta_{t,fixed}$ is the fixed effect slope for time since diagnosis (in years) from the linear mixed effect model (LMEM).

$\hat{\beta}_s$ is calculated as follows:

$$\hat{\beta}_s = D * z^T * (R + z * D * z^T)^{-1} * (Y - x * \hat{\beta}_{fixed})$$

where $D = \begin{bmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{bmatrix}$ is the matrix of variance components of random effects, where V_{11} is the random intercept variance, V_{22} is the random slope variance, and V_{12} is the correlation between the random intercept and random slope. σ_ϵ^2 is the model residual variance, and $R = \sigma_\epsilon^2 * I_{j\ x\ j}$.

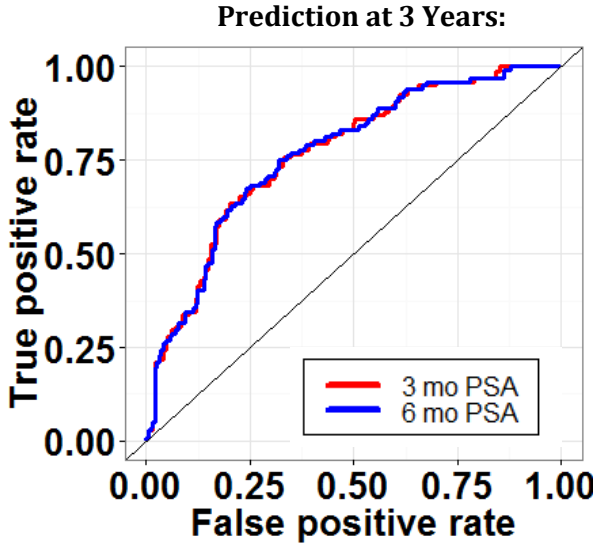
$x = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}$ with fixed effects (intercept and slope) and $z = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}$ with random effects (intercept and slope) contain PSA time points (*t*) up to and including time *j*, and $Y = \begin{bmatrix} PSA_1 \\ \vdots \\ PSA_n \end{bmatrix}$ contains PSA measurements at each time point up to and including time *j*.

In PASS, the model parameters were as follows:

$\beta_{t,fixed} = 0.04$

$D = \begin{bmatrix} 0.39 & 0.002 \\ 0.002 & 0.009 \end{bmatrix}$

$\sigma_\epsilon^2 = 0.07$



Supplemental Figure 1. ROC Curves from Cox PH model with diagnostic PSA plus PSAk using 3 or 6 month PSAs, prediction of reclassification event at 3 years since diagnosis. Prediction was made from 1 year after diagnosis. Cox PH model was adjusted for prostate size, logged time since diagnosis, percent of biopsy cores involved in the most recent biopsy, and history of any negative biopsy after diagnosis. For 3 month PSAs, AUC=0.77, 95% CI=[0.72-0.82]. For 6 month PSAs, AUC=0.77, 95% CI=[0.71-0.82].

Performance of PCA3 and TMPRSS2:ERG urinary biomarkers in prediction of biopsy outcome in the Canary Prostate Active Surveillance Study (PASS)

Lisa F. Newcomb[1,2], Yingye Zheng[3], Anna V. Faino[3], Daniella Bianchi-Frias[4], Matthew R. Cooperberg[5,6], Marshall D. Brown[3], James D. Brooks[7], Peter R. Carroll[5], Atreya Dash[8], Michael D. Fabrizio[9], Martin E. Gleave[10], Michael Liss[11], Todd M. Morgan[12], Ian M. Thompson[13], Andrew A. Wagner[14], Peter S. Nelson[4], and Daniel W. Lin[1,2]

Author affiliations:
[1] Cancer Prevention Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA
[2] Department of Urology, University of Washington, Seattle WA
[3] Biostatistics Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA
[4] Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle WA
[5] Department of Urology, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco CA
[6] Department of Epidemiology & Biostatistics, University of California, San Francisco CA
[7] Department of Urology, Stanford University, Stanford CA
[8] Department of Urology, VA Puget Sound Health Care System, Seattle, WA
[9] Department of Urology, Eastern Virginia Medical School, Virginia Beach VA
[10] Department of Urologic Sciences, University of British Columbia, Vancouver BC
[11] Department of Urology, University of Michigan, Ann Arbor MI
[12] Department of Urology, University of Texas Health Sciences Center, San Antonio TX
[13] CHRISTUS Medical Center Hospital, San Antonio TX
[14] Division of Urology, Beth Israel Deaconess Medical Center, Boston MA

Running title: PCA3 and TMPRSS2:ERG in prostate cancer active surveillance

Keywords: PCA3, TMPRSS2:ERG, biomarkers, prostate cancer, active surveillance

Financial Support: Department of Defense Prostate Cancer Research Program (Grants W81XWH1110489 and W81XWH1410595); National Institutes of Health R01 CA181605; Canary Foundation

Correspondence:
Daniel W. Lin, MD
Department of Urology
University of Washington
1959 NE Pacific Street, Box 356510
Seattle, WA 98195
dlin@uw.edu
phone: 206-221-0797
fax: 206-543-3272

Word count:
Tables and Figures: 6 total (4 tables and 2 figures)
The authors declare no potential conflicts of interest.

ABSTRACT

Purpose:
For men on active surveillance for prostate cancer, biomarkers that are collected non-invasively may improve prediction of reclassification to higher grade or volume cancer. This study examines the association of urinary PCA3 and TMPRSS2:ERG with adverse biopsy reclassification.

Experimental Design:
Urine was collected at baseline, 6, 12, and 24 months as part of the multi-institutional Canary Prostate Active Surveillance Study (PASS) and PCA3 and TMPRSS2:ERG levels were calculated. The association of biomarker scores with reclassification (increase in Gleason score or in ratio of biopsy cores with cancer to ≥34%) was evaluated using multivariable logistic regression models. The association of biomarker scores, or of biomarker kinetics, with time to reclassification was evaluated using multivariable Cox proportional hazards models.

Results:
783 men contributed 2,069 urine specimens. After adjusting for PSA, prostate size and ratio of biopsy cores with cancer, PCA3 was associated with reclassification in 552 men with urine assayed prior to their first surveillance biopsy (HR=1.3; 95% CI 1.0-1.7, p=0.02); no association was found for TMPRSS2:ERG or for either biomarker at subsequent surveillance biopsies. In receiver operating curve analysis, PCA3 did not improve upon the area under the curve of a model with only clinical variables. After adjusting for clinical variables, neither marker was associated with time to reclassification. No association was found for biomarker kinetics and time to reclassification.

Conclusions:
PCA3 associates with cancer reclassification in the first surveillance biopsy, but not at subsequent biopsies. No association is found for urinary TMPRSS2:ERG. Neither marker is likely to change management of patients using active surveillance.

Statement of Translational Relevance:

Active surveillance is the preferred management strategy for many clinically localized prostate cancers. All active surveillance regimens include repeat PSA measurements, clinical exams, and prostate needle biopsies, with curative treatment recommended only if or when higher grade or volume cancer is found. Biomarkers that are collected non-invasively offer the potential for earlier detection of high grade or high volume prostate cancer. In this report, we evaluated the association of urinary PCA3 and TMPRSS2:ERG with reclassification to higher grade or volume cancer in future biopsies. Since there are many different decision points during active surveillance, we assessed if the biomarkers were associated with reclassification at the first or at subsequent surveillance biopsies. We also evaluated if biomarkers collected early in active surveillance associated with time to reclassification, or if changes in biomarkers assayed over time associated with time to reclassification. We found a weak association of PCA3 with reclassification in the first biopsy after diagnosis, but our results suggest that neither of these biomarkers are likely to be useful in changing management of men using active surveillance.

INTRODUCTION

Active surveillance is a management strategy for many clinically localized prostate cancers that allows men to delay or be spared the potential comorbidities of treatment. Cancers that appear low-risk at diagnosis are monitored, typically with regular clinical exams, serial prostate-specific antigen (PSA) measurements, and repeat prostate biopsies. However, uncertainty about harboring occult aggressive cancer and has tempered widespread adoption of active surveillance. Furthermore, optimal surveillance schedules and triggers for intervention have not yet been established, resulting in substantial variation in the practice of active surveillance. Biomarkers that are collected non-invasively and that improve prediction of potentially aggressive tumors could improve utilization of active surveillance and inform decisions about the intensity of surveillance regimens.

During active surveillance, treatment is usually recommended when higher grade or volume disease is found. Biomarkers that detect the presence of occult high grade or high volume disease or that predict future reclassification to high grade or volume cancer could have great utility. There are several clinical scenarios in which a biomarker, or biomarker panel, could demonstrate utility. In all scenarios, a biomarker should incrementally improve upon the performance of commonly available clinical variables.

One question that biomarkers could help answer is if occult high grade cancer is present that was missed in the initial diagnostic biopsy due to the known undersampling of systematic prostate biopsies or uncertainty in emerging magnetic resonance imaging (MRI)-based biopsy protocols. A second question is if biomarkers assessed early in active surveillance improve prediction of time to biopsy reclassification. A third question is if changes in a biomarker over time improve prediction of biopsy reclassification.

Two prostate-specific biomarkers present in urine that showed promise for improving prediction of reclassification in active surveillance include PCA3 and TMPRSS2:ERG mRNA.[1,2] Clinical grade assays have been developed for both of these biomarkers,[2,3] and we have previously shown that the markers were associated with the outcome of baseline biopsy upon entry of 387 men in the multi-center Canary Prostate Active Surveillance Study (PASS).[4] Here, we expanded our analysis to evaluate the utility of the PCA3 and TMPRSS2:ERG assayed at multiple time-points during active surveillance for their association with reclassification to high grade or high volume cancer. We evaluated the performance of biomarkers assayed at a single time-point for association with reclassification at the next biopsy or with time to reclassification at subsequent surveillance biopsies. We also evaluated if biomarkers assayed at several time points during active surveillance changed over time, and if that change associated with time to reclassification.

METHODS

The Canary Prostate Active Surveillance Study (PASS) enrolls men diagnosed with clinically localized prostate cancer who chose to use active surveillance to manage their cancer. Men provide informed consent under institutional review board supervision at nine centers

(clinicaltrials.gov NCT00756665). Under the PASS protocol, PSA is measured every 3 months, clinic visits occur every 6 months, and ultrasound-guided biopsies are performed first between 6 and 12 months after diagnosis, second at 24 months from diagnosis, and then every 2 years. Specimens, including post-DRE urine are collected at study entry and at every 6 month clinic visit. For the present analysis, all urine specimens from study entry, and 6, 12, and 24 month visits that were available in July 2014 were assayed, and follow-up data used for outcomes was frozen February 2017. Men were included in the analysis if they had Gleason score ≤ 3+4 and ≤ 34% ratio of biopsy cores with cancer to total cores collected prior to first urine collection. Men were excluded if they enrolled in PASS > 5 years after diagnosis or had no on-study biopsy for endpoint determination.

Urine samples were collected, assays performed, and biomarker scores calculated as described previously.[4] The natural log of the urine biomarkers was calculated as ln(PCA3) and, due to possible values of 0 in TMPRSS2:ERG, ln(TMPRSS2:ERG + 1).

Statistical methods

Reclassification was defined as an increase in primary or secondary Gleason grade at biopsy and/or an increase in the biopsy cores with cancer to total cores collected (cores ratio) to ≥ 34%. Reclassification in the short term (such as a same day biopsy) and reclassification in the longer term (such as a reclassification 2 years in the future) were evaluated. For all statistical models of interest, a clinical model was first built on routinely collected clinical and biopsy variables including: most recent natural logged PSA, natural logged time since diagnosis (in years), natural logged diagnostic ratio of positive cores, cT-stage (T1a-c versus T2a-c), natural logged prostate size, diagnostic Gleason (3+3 or 3+4), BMI (obese, overweight or normal), race (Caucasian, African American or other), ethnicity (Hispanic versus non-Hispanic/other), age at diagnosis, family history of PCa, alcohol use, and smoking status (current, former or never). Natural logged diagnostic PSA was also considered for the time to reclassification models. For models that incorporated prior biopsy information, maximum prior ratio of positive cores, number of prior biopsies and number of prior negative biopsies were also considered. Insignificant variables were backwards eliminated based on a p-value cutoff of 0.05. Urine biomarkers were secondly added to the models that had been backwards eliminated in order to see if the urine biomarkers were adding information above and beyond variables that were routinely collected.

Binary analysis of reclassification at next biopsy

The association between PCA3 or TMPRSS2:ERG and reclassification at next biopsy was evaluated separately for the first surveillance biopsy and for subsequent surveillance biopsies to allow for model differences. For subsequent surveillance biopsies, participants were included if they had not reclassified on their first surveillance biopsy. The association between urine biomarkers collected prior to biopsy and outcome at the biopsy was modeled using a logistic regression model. A robust variance estimator was used to account for multiple biopsies within participant, where appropriate.

Analysis of time to reclassification for biomarkers assessed after diagnosis and prior to first surveillance biopsy

For time to reclassification, the association between the first PCA3 or TMPRSS2:ERG with time to future reclassification or censoring was modeled using a Cox proportional hazards (PH) model. In order to look at longer term reclassification events, participants were excluded if they reclassified on the first surveillance biopsy. Urine biomarkers closest to diagnosis and prior to first surveillance biopsy were used for the time to reclassification analysis.

Analysis of longitudinal changes in biomarker scores

To assess whether longitudinal urine biomarker data better predicted time to reclassification compared to a single measure, a two-stage procedure was used in the statistical modeling.[5,6] In the first stage, we calculated urine biomarker kinetics based on a linear mixed effect model (LMEM), in which the natural log of the urine biomarkers was modeled as a linear function of time since diagnosis, with a random intercept indicating the individual-specific ln(urine biomarker) at diagnosis, and a random slope reflecting individual-specific rate of change over time. A urine biomarker kinetics value for each participant based on the first urine sample up to a specific observation time was then derived based on the best linear unbiased predictor (BLUP) estimator from the LMEM (reference Cooperberg PSAk paper). Intraclass correlation coefficient (ICC) was calculated in order to assess how much of the variability in the urine biomarkers was explained by between-individual variance compared to total variance. A high ICC indicated strong correlations among urine biomarker measurements from the same individual.

In the second stage, we modeled the risk of future reclassification as a function of covariates at each observation time. The outcome was defined as the time from each urine biomarker measurement to either reclassification or censoring. The first urine biomarker sample along with the urine biomarker kinetics at each observation time were added to the clinical model.

For all time to event analyses, participants without reclassification were censored at date of last study contact, treatment, or 2 years after their last biopsy, whichever came first. All analyses were performed with SAS version 9.4 and R version 3.4.1.

RESULTS

There were 783 participants included in the analysis with a median follow-up time of 4.2 years (IQR: 2.1, 6.1). Median age of participants was 63, median PSA was 4.8, median prostate size was 40, 95% were diagnosed with Gleason 3+3 cancer, and the median ratio of cores containing cancer to total biopsy cores was 8.3% (Table 1). For this analysis, 209 participants contributed 4 urine specimens, 240 contributed 3 specimens, 179 contributed 2 specimens, and 155 contributed 1 urine specimen. Median PCA3 and TMPRSS2:ERG scores were consistently higher in men who reclassify versus those who do not (Figure 1 and Table 1). Figure 1 depicts the design of this study, with results of each analysis described in the sections below.

Binary analysis of reclassification at next biopsy

There were 552 participants with urine assayed prior to their first surveillance biopsy and 230 had their first urine sample collected prior to a subsequent (second, third, etc.) surveillance biopsy; 216 who had their urine collected prior to their first surveillance biopsy also had urine collected prior to subsequent biopsies. The association between PCA3 or TMPRSS2:ERG and reclassification was evaluated separately for the first surveillance biopsy and for subsequent surveillance biopsies.

Of the 552 men with urine biomarkers assessed prior to the first surveillance biopsy, 130 (24%) were adversely reclassified at biopsy, either because they had an increase in Gleason grade or an increase in ratio of cancer-containing to total biopsy cores to ≥ 34%. Men who reclassified at the first biopsy had significantly higher PSA, smaller prostates, and higher cores ratio. They also had a significantly higher PCA3 score, but there was not a significant difference in TMPRSS2:ERG score (Table 1). In a logistic regression model adjusted for PSA, ratio of biopsy cores containing cancer, and prostate size, PCA3 score was significantly associated with reclassification in the first surveillance biopsy (OR = 1.3; 95% CI: 1.0, 1.7, p = 0.02), and TMPRSS2:ERG score was not associated with reclassification (Table 2). However, when receiver operating curve (ROC) analysis was performed using a model with only clinical variables or a model with clinical variables plus PCA3, there was no significant change in the AUC (0.741 [IQR 0.687-0.790] vs 0.751 [0.699-0.796] without and with PCA3 respectively; see Supplementary Figure 1).

In the 446 men with urine biomarkers assessed prior to subsequent surveillance biopsies, 85 (19%) reclassified (Supplementary Table 1). A weaker association was seen between PCA3 and reclassification, but there was a slight association between TMPRSS2:ERG and reclassification. In a logistic regression model adjusted for time since diagnosis, PSA, ratio of biopsy cores containing cancer, prostate size, BMI, and number of biopsies without cancer after diagnosis, neither PCA3 nor TMPRSS2:ERG were associated with reclassification (OR = 1.01; 95% CI: 0.77, 1.32, p = 0.96, and OR = 1.12; 95% CI: 1.00, 1.27, p = 0.06, respectively), but the TMPRSS2:ERG signal was trending towards significant (Table 3).

Analysis of time to reclassification for biomarkers assessed after diagnosis and prior to first surveillance biopsy

There were 405 participants with urine collected prior to the first surveillance biopsy and who did not reclassify or were not censored at the first surveillance biopsy. With a median follow-up of 3.4 years from the first urine collection, 103 (25%) reclassified at a subsequent surveillance biopsy (Supplementary Table 2). In an unadjusted Cox proportional hazards model, PCA3 score was significantly associated with time to reclassification (HR = 1.38; 95% CI: 1.11 – 1.72, p = 0.004), but TMPRSS2:ERG score was not (HR = 1.06; 95% CI: 0.94 – 1.18, p = 0.35). In a model adjusted for prior cores ratio, PSA, and prostate size, the association of PCA3 score and time to reclassification was no longer significant (HR = 1.21; 95% CI: 0.96 – 1.51, p = 0.10), and TMPRSS2:ERG score is still not associated with time to reclassification (HR = 1.05; 95% CI: 0.93 – 1.17, p = 0.43) (Table 4).

Analysis of longitudinal changes in biomarker scores

The annual percent change in PCA3 estimated by the LMEM was 9.8 (95% CI 7.3-12.3, p<0.001). As determined by the ICC, 85% of the observed variation in PCA3 was explained by between-participant variation, and 15% due to within-participant variation. The annual percent change in TMPRSS2:ERG estimated by the LMEM was 11.3 (95% CI 5.2-17.8, p<0.001). As determined by the ICC, 68% of the observed variation in TMPRSS2:ERG was explained by between-participant variation, and 32% due to within-participant variation. No significant differences in slopes were found between event versus long-term non event participants for either PCA3 or TMPRSS2:ERG (Figure 2). In a Cox proportional hazards model adjusted for adjusted for natural logged time since diagnosis, BMI, natural logged prostate size, cores ratio, prior biopsies since diagnosis (0 vs 1+), prior negative biopsies since diagnosis (0 vs 1+), and natural logged recent PSA, there was not a significant association found for the change in PCA3 or TMPRSS2:ERG over time (HR for 0.10 increase in PCA3 = 1.01; 95% CI 0.46 - 2.20, p = 0.98 and in TMPRSS2:ERG = 1.58; 95% CI 0.74, 3.35, p = 0.43).

DISCUSSION

In this report from a multi-center contemporary active surveillance cohort, we evaluated the association between urinary PCA3 and TMPRSS2:ERG and biopsy reclassification. By definition, active surveillance involves monitoring over time and the biomarkers were assessed at multiple time-points. We performed multiple analyses to help inform varying decisions made during active surveillance. Overall, we observed a weak association of PCA3 with reclassification at the first surveillance biopsy (sometimes referred to as the confirmatory biopsy), but no association between PCA3 and time to reclassification. We also did not find evidence that changes in the biomarker scores over time were associated with reclassification. Finally, we did not find evidence that either urinary PCA3 or TMPRSS2:ERG improved association of a model combining readily available clinical variables with reclassification sufficiently to affect clinical management of patients using active surveillance.

In this study we evaluated two urine-based assays in which the specimens were collected non-invasively and the assays were analytically validated. The Progensa PCA3 assay is commercially available and has been FDA approved to inform biopsy decision making in men with no known cancer and a previous negative biopsy; at the time we initiated this work, the TMPRSS2:ERG assay had been analytically validated[2] and was being developed as a commercial assay. PCA3 is a prostate-specific non-coding mRNA and has been shown in many studies to improve predictive accuracy for cancer on initial biopsy,[3,7-9] and to be correlated with more aggressive cancer at prostatectomy.[10,11] Of the genomic alterations involving ETS oncogene family members, a rearrangement involving the androgen-regulated TMPRSS2 gene with the ERG transcription factor (TMPRSS2:ERG) is the most prevalent,[12] occurring in approximately half of the prostate cancers diagnosed in Caucasians[13] and have been correlated in some reports with aggressive disease.[14,15] Thus, we expected that both biomarker assays could improve management of patients using active surveillance.

Biomarkers that improve discrimination of indolent cancers, which will not cause harm if left untreated, and more aggressive tumors, which may benefit from early treatment, will not only support the practice of active surveillance but also promote less intensive biopsy regimes. To optimize patient management and clinical utility, biomarkers should incrementally improve upon existing information. Multimodal risk assessment approaches that combine several sources of information into a risk score have been developed, such as the Prostate Cancer Prevention Trial (PCPT) Risk Calculator, which is used prior to a diagnosis of cancer to predict the risk of finding high grade cancer in the next biopsy.[16] Several commercial biomarker panels employ such a multimodal strategy for predicting the risk of finding high grade cancer prior to a diagnosis of cancer, including the 4Kscore[17] and more recently the Select score.[18] Because men who are candidates for active surveillance have already gone through some risk stratification at their initial biopsy (they have been distinguished from men with no detectable cancer and from men with a finding of high grade cancer), risk models that include information from the index biopsy have been developed for the active surveillance setting.[19,20]

Studies evaluating the use of PCA3 in active surveillance have been limited and results have been conflicting.[4,21,22] In the current study, which is the largest study to date of PCA3 in men using active surveillance, we found a weak but significant association of PCA3 with reclassification at the first surveillance biopsy (adjusted OR = 1.3, p = 0.02), but not for subsequent surveillance biopsies (adjusted OR = 1.01, p = 0.96). Interestingly, when we split our data 2/3 and 1/3 into a training and testing set, we found no significant association of PCA3 and reclassification at the first surveillance biopsy (adjusted OR = 1.2, p = 0.24), indicating that the weak PCA3 signal is unstable. Although we found no association of TMPRSS2:ERG and biopsy reclassification, some studies have suggested improved performance when PCA3 and TMPRSS2:ERG are used in combination[23] or combined into a MiPS score.[24] We thus combined PCA3 and TMPRSS2:ERG into a MIPS score, but found little or no improvement over PCA3 alone (data not shown). Even though PCA3 did associate with reclassification at the first surveillance biopsy, it did not improve prediction of reclassification, as demonstrated by receiver operator characteristic curve: 0.751 vs 0.741 for clinical variables with or without PCA3, respectively.

The analyses described above that have a binary endpoint (reclassification or no reclassification at the next biopsy) may be useful for reducing unnecessary biopsies. We have previously described how patients with a very low risk of reclassifying at the next biopsy, based on models with clinical factors alone or clinical factors and a biomarker panel, could delay or avoid a biopsy.[20] To further help personalize active surveillance regimens, we are also interested in evaluating whether a biomarker, or combination of biomarkers, can improve stratification of men with a very low risk or high risk of having more adverse cancer found in, say, 5 years. Therefore, we evaluated the association of each of the urinary biomarkers, assessed after diagnosis but prior to the first surveillance biopsy, with time to reclassification. Although PCA3 alone was associated with time to reclassification, when incorporated into a multivariate model adjusted for ratio of biopsy cores containing cancer, PSA, and prostate size, neither PCA3 or TMPRSS2:ERG were associated with time to reclassification.

A third way in which biomarkers may improve risk stratification is to incorporate biomarker kinetics into models with clinical variables. Thus, we also evaluated if changes in PCA3 or TMPRSS2:ERG scores measured over time were associated with reclassification. We used samples collected after diagnosis and prior to the first surveillance biopsy, and at 6 month intervals up to 2 years and employed an analytic strategy which allowed the models to account for each individual biomarker measurement while borrowing information from the general trend across all participants and accommodating for random variability in the biomarkers. Although both PCA3 score and TMPRSS2:ERG score were higher in men who reclassified than those who did not, there were not significant differences in the slopes of either biomarker over time (Figure 2) and there was no association of biomarker kinetics with time to reclassification. Our results for PCA3 are consistent with the one other longitudinal study of PCA3 in a smaller, more uniform risk cohort, suggesting that longitudinal PCA3 measurements do not add value over a single PCA3 measurement at any time.[22]

It should be noted that we found a surprisingly large amount of noise in the longitudinal samples, especially for TMPRSS2:ERG. To assess the variation statistically, we used ICC, which can be interpreted as the proportion of total variability explained by between-participant variability. In essence, if a straight line were drawn for each participant depicting the change in biomarker score over time, the ICC indicates how close to straight line each measurement is; 1-ICC is the proportion of variability explained by within-participant variability, or noise. The variation in longitudinal measurements of PCA3 is very similar to that of PSA (submitted for publication); for both markers, 15% of the variability can be attributed to noise. However, 32% of the variability in TMPRSS2:ERG was attributed to noise. The “noise” that we observed in the TMPRSS2:ERG measurements can further be described using the scores of 0. It is known that a third to a half of prostate tumors do not contain the TMPRSS2:ERG gene fusion (ref), and we find that 41% of our participants have at least one TMPRSS2:ERG of 0. However, in the participants with 4 measurements, only 4% have all 4 TMPRS2:ERG scores as 0. If we define 0 as a score less than 20, there are 71% and 36% of participants with 1 or 4, respectively, scores <20. Furthermore, there is a similar distribution of the first, second, third, or fourth score being 0.

This study is not without limitations. First, our design does not allow for us to address if the observed variability is due to biology, specimen collection methods, or assay performance. Second, although this is the largest study of urine biomarkers in active surveillance published to date, the sample size is somewhat small. Another limitation may be the reliability our endpoint. We acknowledge that biopsy reclassification is imperfect in that it may reflect minimal changes in the tumor that may have little clinical importance. None the less, our definition is consistent with those used in most active surveillance cohorts, and importantly does drive treatment decisions in contemporary clinical practice. We did however evaluate biomarker scores in the 31 men who reclassified to 4+3 or higher at their first surveillance biopsy, and found no difference from the scores in the men who reclassified to 3+4 (data not shown).

In conclusion, we found that urinary PCA3 and TMPRSS2:ERG scores are higher in men who have biopsy reclassification while managing their prostate cancer using active surveillance. We found that PCA3 associates with reclassification at the first surveillance biopsy in a

multivariable model, but PCA3, or PCA3 and TMPRSS2:ERG together, did not improve the performance of a multivariable model containing PSA, prostate size and the ration of biopsy cores containing cancer. Furthermore, neither PCA3 or TMPRSS2:ERG, at a single timepoint or measured over time, were associated with time to reclassification. Overall, we conclude that these markers are probably not useful for clinical decision making during active surveillance.

ACKNOWLEDGEMENTS

The authors would like to thank the large and dedicated PASS team, including the coordinators and coordinating center staff. We also thank Scott Tomlins and Jack Groskopf for helpful comments and the research team at Hologic GenProbe for running the biomarker assays. Importantly, the authors also thank all of the men who have participated in PASS.

Figure 1. Study design. Binary analysis of the association of biomarkers with reclassification at the next biopsy was performed for A. 552 men with urine assayed prior to the first surveillance biopsy and B. 446 men with urine assayed prior to subsequent surveillance biopsies (203 with their first study urine prior to subsequent biopsies, and 216 who also had urine collected prior to their first surveillance biopsy). C. The association of biomarkers with time to reclassification was examined in the 405 men with urine assayed prior to their first surveillance biopsy who did not reclassify or were censored at that biopsy. D. The association of biomarker kinetics with time to reclassification was evaluated in the same men as part C.

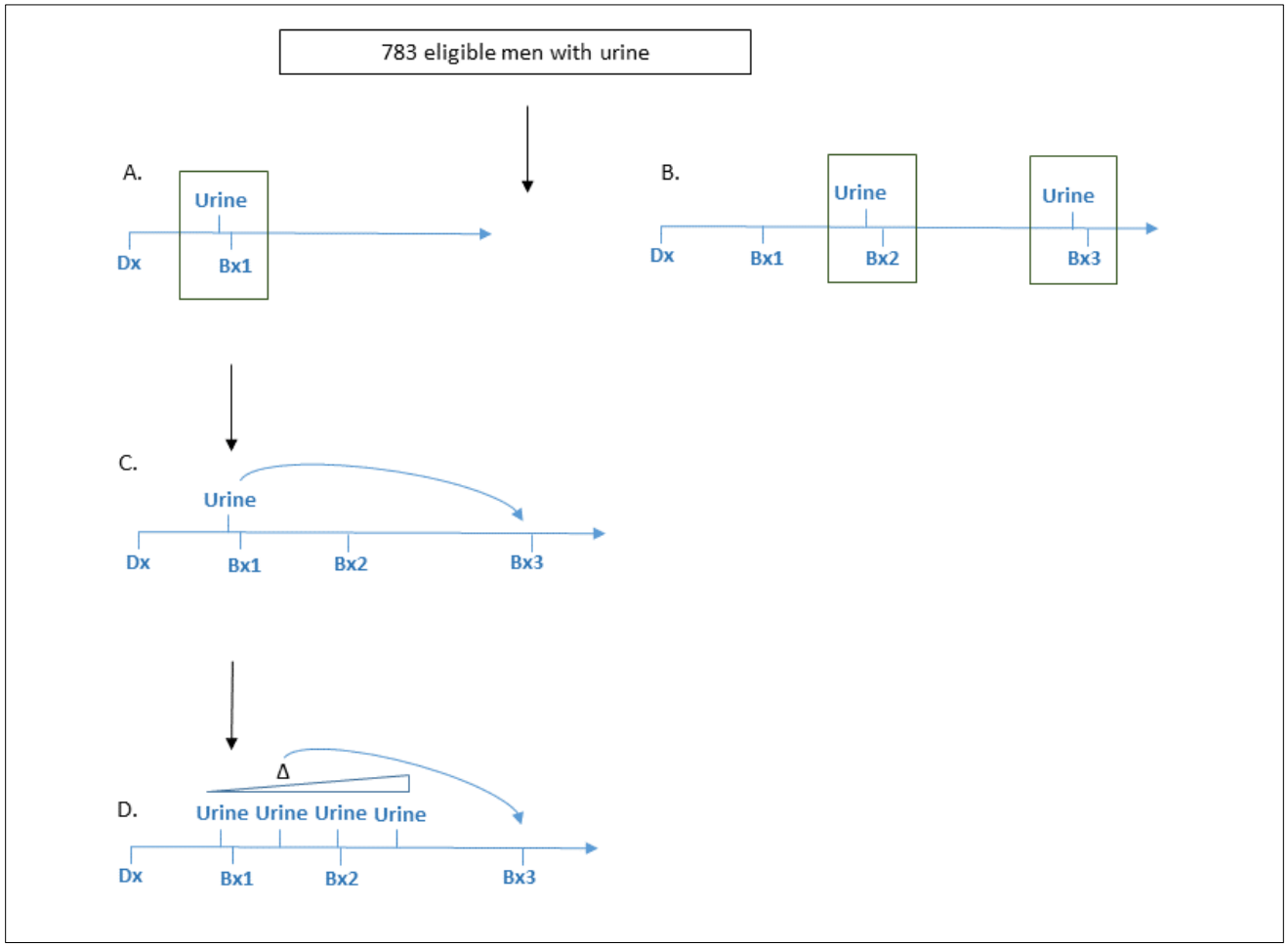


Table 1. Clinical variables for cohort and for participants with urine collected prior to the first surveillance biopsy (Bx1).

Variable	All participants, n=783	Participants with Bx1, n=552	Reclassifiers, n=130	Non-Reclassifiers, n=422	P-value*
	Median [IQR]	Median [IQR]	Median [IQR]	Median [IQR]	
Age at Dx	63 [57, 67]	63 [58, 67]	63 [58, 69]	63 [58, 67]	0.30
Race, n(%)					0.20
Caucasian American	709 (91)	500 (91)	115 (88)	385 (91)	
African American	41 (5)	27 (5)	10 (8)	17 (4)	
Other	33 (4)	25 (5)	5 (4)	20 (5)	
Dx PSA	4.8 [3.6, 6.3]	4.8 [3.8, 6.4]	5.1 [4.3, 6.3]	4.7 [3.6, 6.4]	0.10
Prostate size	40 [29, 55]	40 [30, 55]	35 [25, 48]	42 [32, 60]	<.001
cT-stage, n(%)					0.60
T1a-T1c	707 (90)	508 (92)	118 (91)	390 (92)	
T2a-T2c	76 (10)	44 (8)	12 (9)	32 (8)	
Dx Gleason					0.30
3+3	742 (95)	521 (94)	120 (92)	401 (95)	
3+4	41 (5)	31 (6)	10 (8)	21 (5)	
Dx cores ratio	8.3 [8.3, 16.7]	8.3 [8.3, 16.7]	16.7 [8.3, 24.5]	8.3 [8.3, 16.7]	<.001
BMI, n(%)					0.70
Normal	200 (26)	131 (24)	32 (25)	99 (23)	
Overweight	397 (51)	283 (51)	63 (48)	220 (52)	
Obese	186 (24)	138 (25)	35 (27)	103 (24)	
PCA3	30 [17, 56]	32 [18, 61]	39.5 [24, 89]	30 [16, 57]	<.001
T2:ERG	12 [2, 52]	14 [2, 57]	27 [1, 82]	13 [2, 53]	0.20

* P-value comparing Reclassifiers to Non-Reclassifiers from Wilcoxon rank sum test for continuous variables, and from Fisher's exact test for categorical variables.

Table 2. Association of grade and/or tumor volume reclassification in first surveillance biopsy (n = 552, 130 (24%) with event).

Variable~	Univariable		Multivariable	
	OR (95% CI)^	p-value^	OR (95% CI)^	p-value^
PSA	1.5 (1.1, 2.0)	0.01	1.8 (1.3, 2.6)	0.001
Dx Cores Ratio	4.0 (2.6, 6.3)	<.001	3.4 (2.1, 5.4)	<.001
Prostate size	0.3 (0.2, 0.5)	<.001	0.3 (0.1, 0.4)	<.001
PCA3	1.6 (1.2, 1.9)	0.0001	1.3 (1.0, 1.7)	0.02
T2:ERG	1.1 (1.0, 1.2)	0.21	1.0 (0.9, 1.2)	0.52

^ Odds ratios, 95% confidence intervals and p-values from logistic regression models.

Table 3. Association of grade and/or tumor volume reclassification in subsequent surveillance biopsies (446 participants with 556 biopsies, 85 participants (19%) with event).

Variable	Univariable		Multivariable	
	OR (95% CI)^	p-value^	OR (95% CI)^	p-value^
Time since Dx	0.71 (0.39, 1.29)	0.26	1.37 (0.68, 2.79)	0.38
PSA	1.71 (1.29, 2.27)	<.001	2.28 (1.65, 3.16)	<.001
Cores ratio	1.07 (1.04, 1.10)	<.001	1.04 (1.00, 1.07)	0.04
Prostate size	0.37 (0.23, 0.59)	<.001	0.18 (0.10, 0.33)	<.001
BMI				
Overweight vs Normal	1.36 (0.77, 2.39)	0.29	1.97 (1.07, 3.65)	0.03
Obese vs Normal	1.99 (1.07, 3.68)	0.03	3.31 (1.68, 6.50)	<.001
Negative biopsies after Dx				
1 vs none	0.36 (0.22, 0.60)	<.001	0.51 (0.29, 0.88)	0.02
2 vs none	0.16 (0.06, 0.43)	<.001	0.21 (0.06, 0.69)	0.01
PCA3	1.23 (0.97, 1.55)	0.09	1.01 (0.77, 1.32)	0.96
T2:ERG	1.15 (1.02, 1.28)	0.02	1.12 (1.00, 1.27)	0.06

*Dx = diagnosis; Cores ratio = ratio of biopsy cores containing cancer to total number of cores

~ All variables apart from Cores ratio, BMI, and Negative biopsies after diagnosis were natural log transformed.

^ Odds ratios, 95% confidence intervals and p-values from generalized linear regression models that account for multiple biopsies within each individual.

Table 4. Time to grade and/or tumor volume reclassification using samples collected prior to first surveillance biopsy (n=405, 103 (25%) with event).

Variable~	Univariable		Multivariable	
	HR (95% CI)^	p-value^	HR (95% CI)^	p-value^
cores ratio	2.16 (1.40 – 3.34)	<.001	1.69 (1.08 – 2.64)	0.02
PSA	1.41 (1.02 – 1.95)	0.04	1.84 (1.29 – 2.61)	<.001
Prostate size	0.46 (0.29 – 0.72)	<.001	0.32 (0.19 – 0.54)	<.001
PCA3	1.38 (1.11 – 1.72)	0.004	1.21 (0.96 – 1.51)	0.10
T2:ERG	1.06 (0.95 – 1.19)	0.29	1.05 (0.93 – 1.17)	0.43

~ All variables natural log transformed.

^ Hazard ratios, 95% confidence intervals and p-values from Cox proportional hazards models.

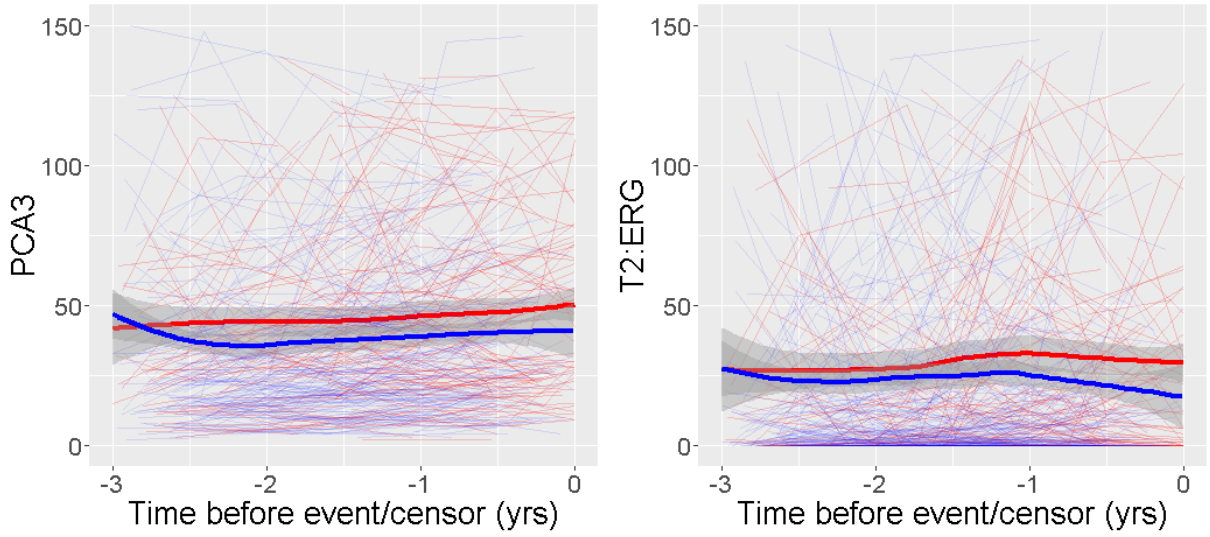


Figure 2. Urine biomarker trajectory prior to event or censor. Event participants in red and censored, non-event participants in blue. Plot does not include urine biomarker data within 2 years prior of censor date in order to look at long-term non-events. Differences in slopes were not significant when assessed with an interaction term in a LMEM (p=0.77 for PCA3, p=0.76 for T2:ERG).

References

1. Bussemakers MJ, van Bokhoven A, Verhaegh GW, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.* 1999;59(23):5975-5979.

2. Tomlins SA, Aubin SM, Siddiqui J, et al. Urine TMPRSS2:ERG Fusion Transcript Stratifies Prostate Cancer Risk in Men with Elevated Serum PSA. *Sci Transl Med.* 2011;3(94):94ra72.

3. Groskopf J, Aubin SM, Deras IL, et al. APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. *Clin Chem.* 2006;52(6):1089-1095.

4. Lin DW, Newcomb LF, Brown EC, et al. Urinary TMPRSS2:ERG and PCA3 in an Active Surveillance Cohort: Results from a Baseline Analysis in the Canary Prostate Active Surveillance Study. *Clin Cancer Res.* 2013;19(9):2442-2450. PMC3674574

5. Zheng YY, Heagerty PJ. Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. *Biostatistics.* 2004;5(4):615-632.

6. Maziarz M, Heagerty P, Cai T, Zheng Y. On longitudinal prediction with time-to-event outcome: Comparison of modeling options. *Biometrics.* 2017;73(1):83-93. PMC5250577

7. Chevli KK, Duff M, Walter P, et al. Urinary PCA3 as a predictor of prostate cancer in a cohort of 3,073 men undergoing initial prostate biopsy. *J Urol.* 2014;191(6):1743-1748.

8. Wei JT, Feng Z, Partin AW, et al. Can urinary PCA3 supplement PSA in the early detection of prostate cancer? *J Clin Oncol.* 2014;32(36):4066-4072. PMC4265117

9. Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. *J Urol.* 2008;179(4):1587-1592.

10. Whitman EJ, Groskopf J, Ali A, et al. PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. *J Urol.* 2008;180(5):1975-1978; discussion 1978-1979.

11. Nakanishi H, Groskopf J, Fritsche HA, et al. PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. *J Urol.* 2008;179(5):1804-1809; discussion 1809-1810.

12. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science.* 2005;310(5748):644-648.

13. Tomlins SA, Bjartell A, Chinnaiyan AM, et al. ETS gene fusions in prostate cancer: from discovery to daily clinical practice. *Eur Urol.* 2009;56(2):275-286.

14. Demichelis F, Fall K, Perner S, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene.* 2007;26(31):4596-4599.

15. Attard G, Clark J, Ambroisine L, et al. Duplication of the fusion of TMPRSS2 to ERG sequence identifies fatal human prostate cancer. *Oncogene.* 2008;27(3):253-263. 2646890

16. Ankerst DP, Hoefler J, Bock S, et al. Prostate Cancer Prevention Trial risk calculator 2.0 for the prediction of low- vs high-grade prostate cancer. *Urology.* 2014;83(6):1362-1367. PMC4035700

17. Parekh DJ, Punnen S, Sjoberg DD, et al. A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. *Eur Urol.* 2015;68(3):464-470.

18. Van Neste L, Hendriks RJ, Dijkstra S, et al. Detection of High-grade Prostate Cancer Using a Urinary Molecular Biomarker-Based Risk Score. *Eur Urol.* 2016.

19. Ankerst DP, Xia J, Thompson IM, Jr., et al. Precision Medicine in Active Surveillance for Prostate Cancer: Development of the Canary-Early Detection Research Network Active Surveillance Biopsy Risk Calculator. *Eur Urol.* 2015;68(6):1083-1088. PMC4583313

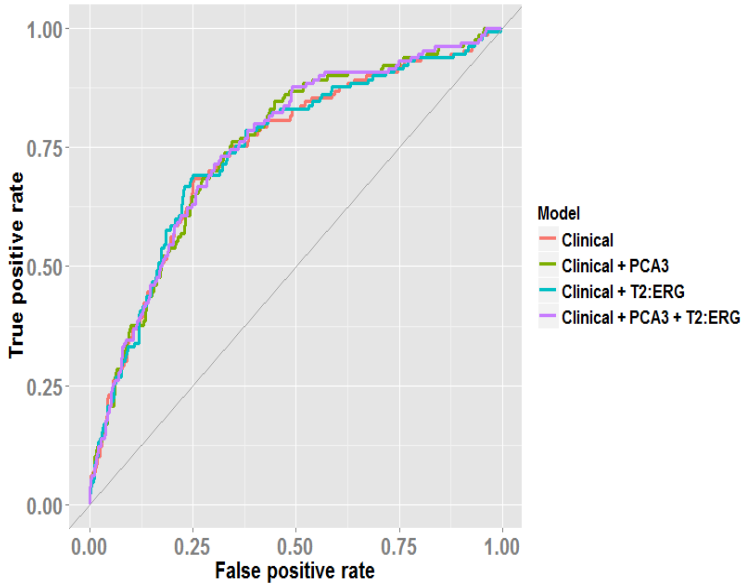
20. Lin DW, Newcomb LF, Brown MD, et al. Evaluating the Four Kallikrein Panel of the 4Kscore for Prediction of High-grade Prostate Cancer in Men in the Canary Prostate Active Surveillance Study. *Eur Urol.* 2016.

21. Tosoian JJ, Loeb S, Kettermann A, et al. Accuracy of PCA3 measurement in predicting short-term biopsy progression in an active surveillance program. *J Urol.* 2010;183(2):534-538.

22. Tosoian JJ, Patel HD, Mamawala M, et al. Longitudinal assessment of urinary PCA3 for predicting prostate cancer grade reclassification in favorable-risk men during active surveillance. *Prostate Cancer Prostatic Dis.* 2017.

23. Sanda MG, Feng Z, Howard DH, et al. Association Between Combined TMPRSS2:ERG and PCA3 RNA Urinary Testing and Detection of Aggressive Prostate Cancer. *JAMA Oncol.* 2017;3(8):1085-1093.

24. Tomlins SA, Day JR, Lonigro RJ, et al. Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. *Eur Urol.* 2015. PMC4644724



Model	AUC (95% CI)
Clinical variables alone (No urine)	0.741 (0.687 – 0.790)
Clinical + PCA3	0.751 (0.699 – 0.796)
Clinical + T2:ERG	0.742 (0.688 – 0.794)
Clinical + PCA3 + T2:ERG	0.752 (0.701 – 0.799)

Supplementary Fig 1. Validation of model predicting reclassification at the first surveillance biopsy (Bx1).

Supplemental Table 1. Clinical variables for participants with urine collected prior to subsequent surveillance biopsies (i.e. the second or third surveillance biopsy).

Variable	Participants with Subsequent Bx, n=446	Reclassifiers, n=85	Non-Reclassifiers, n=361	P-value*
	Median [IQR]	Median [IQR]	Median [IQR]	
Age at Dx	62 [57, 66]	63 [58, 66]	61 [57, 66]	0.30
Race, n(%)				0.40
Caucasian American	409 (92)	78 (92)	331 (92)	
African American	19 (4)	2 (2)	17 (5)	
Other	18 (4)	5 (6)	13 (4)	
Dx PSA	4.6 [3.4, 6.1]	4.8 [3.6, 6.2]	4.5 [3.3, 6.1]	0.30
Prostate size	41 [30, 57]	35 [25, 45]	43 [31, 60]	<.001
cT-stage, n(%)				0.40
T1a-T1c	405 (91)	75 (88)	330 (91)	
T2a-T2c	41 (9)	10 (12)	31 (9)	
Dx Gleason				0.10
3+3	429 (96)	79 (93)	350 (97)	
3+4	17 (4)	6 (7)	11 (3)	
Dx cores ratio	8.3 [8.3, 16.7]	14.3 [8.3, 20.0]	8.3 [8.3, 15.1]	<.001
BMI, n(%)				0.20
Normal	125 (28)	19 (22)	106 (29)	
Overweight	221 (50)	42 (49)	179 (50)	
Obese	100 (22)	24 (28)	76 (21)	
PCA3^	32 [18, 58]	43 [23, 66]	31 [17, 56]	0.02
T2:ERG^	16 [2, 61]	32 [5, 82]	13 [1, 57]	0.02

^ Urine biomarker median [IQR] from earliest subsequent biopsy.

* P-value comparing Reclassifiers to Non-Reclassifiers from Wilcoxon rank sum test for continuous variables, and from Fisher's exact test for categorical variables.

Supplemental Table 2. Clinical variables for participants with urine collected prior to the first surveillance biopsy that are included in the time-to-event analysis.

Variable	Participants with urine for time to event,^ n=405	Reclassifiers, n=103	Non-Reclassifiers, n=302	P-value*
	Median [IQR]	Median [IQR]	Median [IQR]	
Age at Dx	63 [58, 67]	63 [59, 66]	63 [57, 67]	0.99
Race, n(%)				0.70
Caucasian American	370 (91)	93 (90)	277 (92)	
African American	16 (4)	4 (4)	12 (4)	
Other	19 (5)	6 (6)	13 (4)	
Dx PSA	4.8 [3.6, 6.4]	4.7 [3.5, 5.9]	4.8 [3.7, 6.6]	0.30
Prostate size	43 [32, 60]	38 [28, 50]	45 [34, 62]	<.001
cT-stage, n(%)				0.99
T1a-T1c	375 (93)	96 (93)	279 (92)	
T2a-T2c	30 (7)	7 (7)	23 (8)	
Dx Gleason				0.60
3+3	385 (95)	97 (94)	288 (95)	
3+4	20 (5)	6 (6)	14 (5)	
Dx cores ratio	8.3 [8.3, 16.7]	8.3 [8.3, 16.7]	8.3 [8.3, 16.7]	0.01
BMI, n(%)				0.30
Normal	96 (24)	23 (22)	73 (24)	
Overweight	213 (53)	50 (49)	163 (54)	
Obese	96 (24)	30 (29)	66 (22)	
PCA3	28 [16, 51]	30 [20, 56]	27 [14, 48]	0.05
T2:ERG	11 [1, 49]	14 [1, 77]	11 [2, 41]	0.40

^ Participants were required to have urine sample prior to first surveillance biopsy, and could not reclassify on first surveillance biopsy.

* P-value comparing Reclassifiers to Non-Reclassifiers from Wilcoxon rank sum test for continuous variables, and from Fisher's exact test for categorical variables.

The Role of Surveillance Biopsy with No Cancer as a Prognostic Marker for Reclassification: Results from the Canary Prostate Active Surveillance Study (PASS)

James T Kearns[1], Anna V Faino[2], Lisa F Newcomb[1,2], James D Brooks[3], Peter R Carroll[4], Atreya Dash[1], William J Ellis[1], Michael Fabrizio[5], Martin E Gleave[6], Todd M Morgan[7], Peter S Nelson[2], Ian M Thompson[8], Andrew A Wagner[9], Yingye Zheng[2], Daniel W Lin[1]

[1]Department of Urology, University of Washington
[2]Fred Hutchinson Cancer Research Center
[3]Stanford University
[4]University of California San Francisco
[5]Eastern Virginia Medical School
[6]University of British Columbia
[7]University of Michigan
[8]University of Texas Health Sciences Center at San Antonio
[9]Beth Israel Deaconess Medical Center

Source of Funding: Canary Foundation, Institute for Prostate Cancer Research, and the Office of the Assistant Secretary of Defense for Health Affairs, through the Prostate Cancer Research Program under Award No. W81XWH-14-1-0595.

Corresponding Author:
James T Kearns
Department of Urology
University of Washington School of Medicine
1959 NE Pacific, Box 356510
Seattle, WA 98195
Phone: (864) 650-5255 Fax: 206-543-3272
Email: jkearns1@uw.edu

Word count: 2395 words

Portions of this work were presented as a moderated poster at the AUA Annual Meeting, May 2017.

Abstract

Background
Many patients who are on active surveillance (AS) for prostate cancer will have surveillance prostate needle biopsies (PNB) without any cancer evident.

Objective
To define the association between negative surveillance PNB and risk of reclassification on AS.

Design, Setting, and Participants
All men were enrolled in the Canary Prostate Active Surveillance Study (PASS) between 2008-2016. Men were included if they had Gleason ≤ 3+4 prostate cancer and < 34% core involvement ratio at diagnosis. Men were prescribed surveillance PNB at 12 and 24 months after diagnosis and then every 24 months.

Outcome Measurements and Statistical Analysis
Reclassification was defined as an increase in Gleason grade and/or an increase in the ratio of biopsy cores with cancer to ≥ 34%. PNB outcomes were defined as: a) no cancer on biopsy, b) cancer without reclassification, or c) reclassification. Kaplan-Meier and Cox proportional hazards models were performed to assess risk of reclassification.

Results and Limitations
657 men met inclusion criteria. On first surveillance PNB, 214 (32%) had no cancer, 282 (43%) had cancer but no reclassification, and 161 (25%) reclassified. Among those who did not reclassify, 313 had a 2nd PNB. On 2nd PNB, 120 (38%) had no cancer, 139 (44%) had cancer but no reclassification, and 54 (17%) reclassified. In a multivariable analysis, significant predictors of decreased future reclassification after 1st PNB were no cancer on PNB (HR = 0.50, p = 0.008), lower serum PSA, larger prostate size, and lower BMI. A finding of no cancer on the 2nd PNB was also associated with significantly decreased future reclassification in a multivariable analysis (HR = 0.15, p = 0.003), regardless of first PNB result. The major limitation of this study is relatively small numbers of patients with long term follow up.

Conclusions
Men who have a surveillance PNB with no evidence of cancer are significantly less likely to reclassify on AS in the PASS cohort. These finding have implications for tailoring AS protocols.

Patient Summary
Men on active surveillance for prostate cancer who have a biopsy showing no cancer are at decreased risk of having worse disease in the future. This may have an impact on how frequently biopsies need to be performed in the future.

Introduction

Active surveillance (AS) for prostate cancer is an increasingly popular management strategy for Gleason 3+3 and low volume 3+4 prostate cancer.[1] Patients are generally assessed by periodic serum prostate specific antigen (PSA) testing, digital rectal examination, and prostate biopsy. Despite increasing use, an optimal AS protocol that defines precise timing of these assessments has not yet been established or defined by practice guidelines. In published series, biopsies are performed as frequently as annually[2] to every 3-4 years.[3] Furthermore, within a given protocol, there has been no formal strategy for tailoring biopsy frequency based on a patient's individualized risk.

Prostate biopsies yield a wealth of information about an individual's cancer, but many men find them to be unpleasant, they are costly,[4] and there is an approximately 5% risk of infection following biopsy.[5] Furthermore, published active surveillance series report that although the majority of surveillance biopsies find no change in the Gleason grade, 21-50%[6] of surveillance biopsies have no cancer found on the biopsy specimens, suggesting low cancer volume. Given these considerations, it is a common clinical scenario for an AS patient who has one or more surveillance biopsies with the finding of no cancer to question the need for further biopsy.

In this context, we examined the predictive value of no cancer on surveillance biopsy for future pathologic reclassification after a diagnosis of very low- and low-risk prostate cancer in the large, multicenter Canary Prostate Acitve Surveillance Study (PASS). We assessed the significance of biopsy results in the first and second biopsies after the initial diagnosis and performed modeling to take into account variables that contribute to risk of reclassification.

Patients and Methods

Patient Population

PASS is a multi-institutional prostate cancer active surveillance cohort in North America.[7] All patients were enrolled in PASS, approved by institutional review boards at all participating sites (clinicaltrials.gov NCT000756665). Under the PASS protocol, PSA is measured every 3 months, clinic visits occur every 6 months, and ultrasound-guided biopsies are performed first between 6 and 12 months after diagnosis, 24 month after diagnosis and then every 2 years. At least 10-core templates were required, with the median [IQR] number of total biopsy cores collected being 12 [12, 14]. Other tests, including magnetic resonance imaging, may be performed at the clinicians' discretion, but as the study started enrollment in 2008, the majority of men have not undergone these procedures. Patients were included in the current analysis if they were enrolled as of February 2016, had Gleason ≤ 3+4 prostate cancer, had < 34% ratio of biopsy cores containing cancer to total biopsy cores (cores ratio) at diagnosis, and had their first surveillance biopsy after the initial diagnosis of prostate cancer (a.k.a. confirmatory biopsy) within 2 years of diagnosis and while enrolled in PASS.

Outcomes and Statistical Methods

The primary outcome was time to reclassification from either the first or second surveillance biposy. Reclassification was defined as an increase in primary or secondary Gleason grade at biopsy and/or an increase in the cores ratio to ≥ 34%. All pathology outcomes were determined by uropathologists at each site. Sensitivity analyses including participants diagnosed with Gleason 3+3 only or for grade-only reclassification were also performed. Patients without

reclassification were censored at date of last study contact, treatment, or 2 years after their last biopsy, whichever came first.

Patients were stratified by the outcome of their first or second surveillance biopsy as follows: a) no evidence of cancer on biopsy, b) evidence of cancer on biopsy without reclassification, or c) reclassification. Kaplan-Meier curves were plotted to examine how reclassification-free probability varied by surveillance biopsy outcome over the follow up period. Log-rank tests were used to compare differences in reclassification-free probabilities.

Associations between previous surveillance biopsy result (no cancer versus cancer without reclassification) and time to future reclassification were modeled using Cox proportional hazards models. In order to assess whether the first surveillance biopsy result was associated with future reclassification, we considered a time since first surveillance biopsy model, where the association of interest was the result of the first surveillance biopsy. In order to assess whether the aggregate effect of the first and second surveillance biopsy results was associated with future reclassification, we considered a time since second surveillance biopsy model, where the two associations of interest were the results of the first and second surveillance biopsies, respectively. Due to our hypotheses of interest, previous surveillance biopsy result(s) remained in the two models regardless of statistical significance. In addition, the following covariates were considered: natural log transformed PSA closest and prior to surveillance biopsy, maximum cores ratio from either diagnostic biopsy or surveillance biopsy, natural log transformed diagnostic PSA, body mass index (BMI), natural log transformed prostate volume, age at diagnosis, clinical T stage (T1 versus T2), diagnostic Gleason (3+4 or 3+3), and race (Caucasian

versus other). Study site was accounted for by stratifying the baseline hazard. In order to account for potential collinearity among the variables, insignificant covariates were backwards eliminated based on a p-value cutoff of 0.05. Analyses were performed with SAS version 9.4 and R version 3.3.0.

Results

Six hundred fifty-seven men were included in this analysis. Overall median follow up from diagnosis for participants without a reclassification event was 2.9 years (IQR 1.8 – 4.7). All participants received a first surveillance biopsy, which occurred at a median of 1.0 years after diagnosis (IQR 0.7 – 1.2 years). The outcomes of the first surveillance biopsy were 214 (32%) with no cancer on this biopsy, 282 (43%) with cancer on biopsy but no reclassification, and 161 (25%) with reclassification (Figure 1). Of the 496 men who did not reclassify, 313 had a second biopsy at a median of 2.3 years from diagnosis (IQR 2.0 – 3.0 years). Among these 313 men, 120 (38%) had no cancer on this biopsy, 139 (45%) had some cancer but no reclassification, and 54 (17%) had a reclassification event at second biopsy (Figure 1).

The mean age of the cohort was 63, median PSA was 4.9 ng/ml, median prostate volume was 42 cc, 94% were diagnosed with Gleason 3+3, and the median cores ratio was 8% (which corresponds to 1/12 biopsy cores with cancer; Table 1). When stratified by the outcome of the first surveillance biopsy, the groups were similar with respect to racial makeup, age, clinical stage, family history of prostate cancer, and BMI. There were statistically significant differences across groups for prostate volume, serum PSA level, PSA density, diagnostic Gleason grade, and

diagnostic cores ratio positive for prostate cancer (Table 1). The results for patients who underwent a second surveillance biopsy are similar and are given in Supplemental Table 1.

Kaplan-Meier analysis of reclassification stratified by outcome of the first surveillance biopsy is shown in Figure 2. There was a statistically significant difference in time to reclassification in men whose first biopsy had no evidence of cancer versus evidence of cancer without reclassification ($p < 0.001$). Similarly, there was a statistically significant difference in time to reclassification based on the outcome of the second biopsy ($p < 0.001$), as shown in Figure 3. When patients who had two surveillance biopsies without reclassification were stratified by outcome of both first and second surveillance biopsy, the reclassification-free probability was similar for patients whose second surveillance biopsy showed no cancer, regardless of the result of the first biopsy (Supplemental Figure 1).

A first surveillance prostate biopsy negative for any cancer versus positive for cancer without reclassification was associated with less risk of reclassification in future biopsies (HR = 0.44, $p < 0.001$). After adjusting for serum PSA, prostate volume, and BMI, no cancer on initial surveillance biopsy was still significantly protective against reclassification (HR 0.50, $p = 0.008$) (Table 2). Finding no cancer in the second surveillance biopsy was also significantly protective against reclassification in both an unadjusted (HR 0.12, $p < 0.001$) and adjusted (HR 0.18, $p = 0.01$) analysis (Table 3).

All results were similar when sensitivity analysis was performed for grade-only reclassification or for the subset of participants diagnosed with Gleason 3+3 cancer and can be found in the supplemental materials.

Discussion

Our present study examined the risk of pathological reclassification in AS patients who have no cancer on first or second surveillance biopsy. In both Kaplan-Meier and multivariable-adjusted Cox proportional hazard analyses, no cancer on surveillance biopsy was prognostic against future reclassification. When there was no detectable cancer in the first surveillance biopsy, the risk of future reclassification was decreased by 50%, and if no cancer was seen on second surveillance biopsy, then there was an 82% decreased risk of future reclassification.

We also found that patients with no cancer on first surveillance biopsy were more likely to have no cancer on the second surveillance biopsy when compared to those who had a first surveillance biopsy with cancer but no reclassification. This is consistent with previous work suggesting that no cancer found on initial surveillance biopsy is protective against future reclassification[8–11] and work suggesting that negative biopsy prior to diagnosis is associated with lower adverse pathological outcomes at RP.[12] Importantly, it also appears that continued presence of cancer on subsequent surveillance biopsy results in a significantly higher risk of pathological reclassification. Within 5 years of diagnosis, ~3-5% of patients with no cancer on surveillance biopsies reclassify versus ~20-30% of those who have some cancer on subsequent biopsies. These findings indicate that even in men who do not initially reclassify, there is a persistent risk of pathological reclassification and thus a need for continued surveillance. Decreasing risk of

reclassification with increasing biopsy number was seen in this cohort, with 25% of men reclassifying on first biopsy, and 17% reclassifying on second biopsy. This is consistent with our previously reported data and other AS cohorts that demonstrate decreasing rates of reclassification over time.[3,7,13–15]

One of the major goals of evaluating factors that predict reclassification of prostate cancer on AS is to use all available data in the best possible manner to decrease the number of prostate biopsies required without sacrificing the detection of potentially lethal prostate cancer. Laviana et al found that the economic cost of AS increases steadily with time, surpassing the cost of brachytherapy within 9 years and nearly equaling that of RALP by 12 years.[4] These costs were driven chiefly by serial prostate biopsy. In addition to the financial cost of biopsies, there are biopsy-related morbidities, most notably an approximately 5% risk of infection[5]. However, as seen in the ProtecT trial, a strategy of "active monitoring" that relies solely upon large increases in serum PSA levels to trigger prostate biopsy may be an inadequate paradigm, with a 2.6 times increased risk of clinical progression.[16] One or more mandatory surveillance biopsies are likely necessary to better risk stratify patients before making decisions regarding future biopsy frequency. Using a finding that is prognostic against reclassification, such as surveillance biopsy without cancer, to decrease biopsy frequency may decrease patient discomfort, cost, and risk of infection while maintaining detection of significant disease.

In order to best use available clinical information, it is worth noting that risk of reclassification associated with a given variable changes depending on what has transpired with the patient during his course of surveillance. Previously published nomograms for reclassification while on

AS[17,18] do not adjust their covariates over the course of AS, despite patients having different risk profiles as they undergo biopsies without reclassification. We found that no cancer on second surveillance biopsy was much more prognostic against reclassification than no cancer on the first surveillance biopsy (HR 0.18 vs 0.50). This finding is consistent with previous reported outcomes where fewer men reclassify on AS over time.[10] Given that clinical variables may confer different risk at different time points, models and risk assessment tools should account for these varying risks.

Major strengths of our study include the fact that it is a multicenter, prospectively designed study with quality control of all clinical data collected. All participants were recommended the same biopsy schedule (6-12 months after diagnosis, 24 months after diagnosis, and then every 2 years), regardless of whether or not they had detectable disease on surveillance biopsies. Overall, 80% of biopsies were per protocol (on-time), and finding no cancer in the first surveillance biopsy was not associated with delayed subsequent biopsies. The inclusion at diagnosis of both Gleason 3+3 and 3+4 disease makes the results more generalizable to community AS protocols. In addition, the use of pathologic reclassification as the endpoint does not rely upon patient factors such as tolerance for risk or anxiety that may sway treatment decisions. The study is limited by lack of centralized pathologic review, lack of information for all patients regarding magnetic resonance imaging (MRI) use in the surveillance of these men, and relatively small numbers of patients with long term follow up. These limitations are mitigated by the fact that early central pathology review indicates ~80% concordance with local pathology scoring, and most patients in PASS have not had prostatic MRI. Additionally, MRI is still not considered standard of care in AS per National Comprehensive Cancer Network guidelines.[19] Inclusion of

more patients over time with similar risk profiles would be expected to tighten the confidence intervals rather than significantly change hazard ratios. In addition, our study would benefit from validation by an external AS cohort.

Conclusions

No detectable cancer in a biopsy during active surveillance was prognostic for decreased risk of pathologic reclassification. The clinical impact of no cancer on surveillance biopsy becomes stronger on subsequent biopsy, suggesting that risk of reclassification changes with time. Men with Gleason 3+3 PCa and two initial surveillance biopsies with no detectable cancer may not warrant annual or semi-annual biopsy and perhaps may lengthen biopsy interval to several years similar to other published protocols.[3] Further work with models should include the concept of varying risk by taking into account real time variables along the course of AS in order to individualize biopsy intervals and patient assessments.

References

[1] Cooperberg MR, Carroll PR. Trends in Management for Patients With Localized Prostate Cancer, 1990-2013. JAMA 2015;314:80–2. doi:10.1001/jama.2015.6036.

[2] Tosoian JJ, Mamawala M, Epstein JI, Landis P, Wolf S, Trock BJ, et al. Intermediate and Longer-Term Outcomes From a Prospective Active-Surveillance Program for Favorable-Risk Prostate Cancer. J Clin Oncol 2015;33:3379–85. doi:10.1200/JCO.2015.62.5764.

[3] Klotz L, Vesprini D, Sethukavalan P, Jethava V, Zhang L, Jain S, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol Off J Am Soc Clin Oncol 2015;33:272–7. doi:10.1200/JCO.2014.55.1192.

[4] Laviana AA, Ilg AM, Veruttipong D, Tan H-J, Burke MA, Niedzwiecki DR, et al. Utilizing time-driven activity-based costing to understand the short- and long-term costs of treating localized, low-risk prostate cancer. Cancer 2016;122:447–55. doi:10.1002/cncr.29743.

[5] Wagenlehner FME, van Oostrum E, Tenke P, Tandogdu Z, Çek M, Grabe M, et al. Infective complications after prostate biopsy: outcome of the Global Prevalence Study of Infections in Urology (GPIU) 2010 and 2011, a prospective multinational multicentre prostate biopsy study. Eur Urol 2013;63:521–7. doi:10.1016/j.eururo.2012.06.003.

[6] Dall'Era MA, Albertsen PC, Bangma C, Carroll PR, Carter HB, Cooperberg MR, et al. Active Surveillance for Prostate Cancer: A Systematic Review of the Literature. Eur Urol 2012;62:976–83. doi:10.1016/j.eururo.2012.05.072.

[7] Newcomb LF, Thompson IM, Boyer HD, Brooks JD, Carroll PR, Cooperberg MR, et al. Outcomes of Active Surveillance for Clinically Localized Prostate Cancer in the Prospective, Multi-Institutional Canary PASS Cohort. J Urol 2016;195:313–20. doi:10.1016/j.juro.2015.08.087.

[8] Eggener SE, Mueller A, Berglund RK, Ayyathurai R, Soloway C, Soloway MS, et al. A Multi-Institutional Evaluation of Active Surveillance for Low-Risk Prostate Cancer. J Urol 2009;181:1635–41. doi:10.1016/j.juro.2008.11.109.

[9] Tseng KS, Landis P, Epstein JI, Trock BJ, Carter HB. Risk Stratification of Men Choosing Surveillance for Low Risk Prostate Cancer. J Urol 2010;183:1779–85. doi:10.1016/j.juro.2010.01.001.

[10] Porten SP, Whitson JM, Cowan JE, Cooperberg MR, Shinohara K, Perez N, et al. Changes in Prostate Cancer Grade on Serial Biopsy in Men Undergoing Active Surveillance. J Clin Oncol 2011;29:2795–800. doi:10.1200/JCO.2010.33.0134.

[11] Cary KC, Cowan JE, Sanford M, Shinohara K, Perez N, Chan JM, et al. Predictors of Pathologic Progression on Biopsy Among Men on Active Surveillance for Localized Prostate Cancer: The Value of the Pattern of Surveillance Biopsies. Eur Urol 2014;66:337–42. doi:10.1016/j.eururo.2013.08.060.

[12] ElShafei A, Nyame Y, Kara O, Badawy A, Amujiogu I, Fareed K, et al. More Favorable Pathological Outcomes in Men with Low Risk Prostate Cancer Diagnosed on Repeat versus Initial Transrectal Ultrasound Guided Prostate Biopsy. J Urol 2016;195:1767–72. doi:10.1016/j.juro.2015.12.079.

[13] Tosoian JJ, Trock BJ, Landis P, Feng Z, Epstein JI, Partin AW, et al. Active Surveillance Program for Prostate Cancer: An Update of the Johns Hopkins Experience. J Clin Oncol 2011;29:2185–90. doi:10.1200/JCO.2010.32.8112.

[14] Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. Eur Urol 2013;63:597–603. doi:10.1016/j.eururo.2012.11.005.
[15] Dall'Era MA, Konety BR, Cowan JE, Shinohara K, Stauf F, Cooperberg MR, et al. Active surveillance for the management of prostate cancer in a contemporary cohort. Cancer 2008;112:2664–70. doi:10.1002/cncr.23502.
[16] Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med 2016. doi:10.1056/NEJMoa1606220.
[17] Ankerst DP, Xia J, Thompson Jr. IM, Hoefler J, Newcomb LF, Brooks JD, et al. Precision Medicine in Active Surveillance for Prostate Cancer: Development of the Canary–Early Detection Research Network Active Surveillance Biopsy Risk Calculator. Eur Urol 2015;68:1083–8. doi:10.1016/j.eururo.2015.03.023.
[18] Mamawala MM, Rao K, Landis P, Epstein JI, Trock BJ, Tosoian JJ, et al. Risk prediction tool for grade re-classification in men with favourable-risk prostate cancer on active surveillance. BJU Int 2016:n/a-n/a. doi:10.1111/bju.13608.
[19] Prostate Cancer (Version 2.2017. Natl Compr Cancer Netw n.d. https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf (accessed June 26, 2017).

Figure Legends

Figure 1: Consort diagram of patients receiving surveillance biopsy and biopsy outcomes

Bx1 – first surveillance biopsy

Bx2 – second surveillance biopsy

Figure 2: Time to grade and/or tumor volume reclassification by first surveillance biopsy outcome

Bx1 – first surveillance biopsy

Bx1- – no cancer detected on first surveillance biopsy

Bx1+ – cancer but no reclassification detected on first surveillance biopsy

Figure 3: Time to grade and/or tumor volume reclassification by second surveillance biopsy outcome

Bx2 – second surveillance biopsy

Bx2- – no cancer detected on second surveillance biopsy

Bx2+ – cancer but no reclassification detected on second surveillance biopsy

Tables

Table 1: Patient characteristics based on results of first surveillance biopsy

	All Patients	No Cancer 1st Surveillance Biopsy	Cancer without Reclassification 1st Surveillance Biopsy	Reclassification on 1st Surveillance Biopsy	P-value^
N	657	214	282	161	
Race, n(%)					0.12
Caucasian American	588 (89)	187 (87)	258 (91)	143 (89)	
African American	38 (6)	16 (7)	9 (3)	13 (8)	
other	31 (5)	11 (5)	15 (5)	5 (3)	
Prostate volume, cc median [IQR]	42 [31 - 57]	46 [34 - 64]	43 [32 - 56]	36 [27 - 48]	< 0.001
Age, years mean (SD)	63 (7)	62 (7)	63 (7)	63 (7)	0.22
PSA, ng/ml median [IQR]	4.9 [3.9 - 6.5]	5.1 [3.7 - 6.6]	4.7 [3.7 - 6.1]	5.3 [4.4 - 6.6]	0.02
PSA density, median [IQR]	0.11 [0.08 - 0.16]	0.10 [0.07 - 0.14]	0.11 [0.08 - 0.15]	0.15 [0.11 - 0.21]	< 0.001
Clinical stage, n(%)					0.37
T1a-T1c	592 (90)	197 (92)	249 (88)	146 (91)	
T2a-T2c	65 (10)	17 (8)	33 (12)	15 (9)	
Diagnostic Gleason score, n(%)					0.03
3+3	615 (94)	208 (97)	259 (92)	148 (92)	
3+4	42 (6)	6 (3)	23 (8)	13 (8)	
Diagnostic cores ratio, median [IQR]~	8 [8 - 17]	8 [8 - 14]	13 [8 - 17]	17 [8 - 18]	< 0.001
Family history of prostate cancer, n(%)~	176 (28)	55 (27)	79 (29)	42 (27)	0.89
BMI, mean (SD)	28.0 (4.4)	28.2 (4.3)	27.6 (4.0)	28.4 (5.0)	0.08

^ p-value comparing biopsy outcome from 1st surveillance biopsy (no cancer, cancer without reclassification, or reclassification) from Chi-squared test for categorical variables and ANOVA for continuous variables. For prostate volume, PSA, PSA density,and cores ratio, p-value from Kruskal-Wallis test.

~ Cores ratio missing for 38 participants and family history of prostate cancer missing for 21 participants.

Table 2: Time to grade and/or tumor volume reclassification, from time of first surveillance biopsy (n = 494, 85 with event)*

Variable	Univariable		Multivariable	
	HR (95% CI)^	**P-value^**	**HR (95% CI)^**	**P-value^**
No cancer on first surveillance biopsy (versus cancer without reclassification)	0.44 (0.27, 0.71)	< 0.001	0.50 (0.30, 0.83)	0.008
Ln(PSA on/prior to first surveillance biopsy)	1.93 (1.32, 2.83)	< 0.001	2.74 (1.83, 4.10)	< 0.001
Ln(Prostate volume, cc)	0.38 (0.22, 0.69)	0.001	0.19 (0.10, 0.37)	< 0.001
BMI	1.03 (0.98, 1.09)	0.28	1.07 (1.01, 1.13)	0.02

* Two participants were missing cores ratio data and were not included in the modeling.
^ 95% confidence intervals and p-values from Cox proportional hazards models.

Table 3: Time to grade and/or tumor volume reclassification, from time of second surveillance biopsy (n = 259, 29 with event)

Variable	Univariable		Multivariable	
	HR (95% CI)^	**P-value^**	**HR (95% CI)^**	**P-value^**
No cancer on second surveillance biopsy (versus cancer without reclassification)	0.12 (0.03, 0.39)	< 0.001	0.18 (0.05, 0.66)	0.01
No cancer on first surveillance biopsy (versus cancer without reclassification)	0.35 (0.15, 0.82)	0.02	0.53 (0.20, 1.41)	0.20
Ln(PSA on/prior to second surveillance biopsy)	4.66 (2.22, 9.78)	< 0.001	6.10 (2.62, 14.17)	< 0.001
Ln(prostate volume, cc)	0.45 (0.16, 1.26)	0.13	0.18 (0.05, 0.64)	0.008

^ 95% confidence intervals and p-values from Cox proportional hazards models.

Figures:

Figure 1

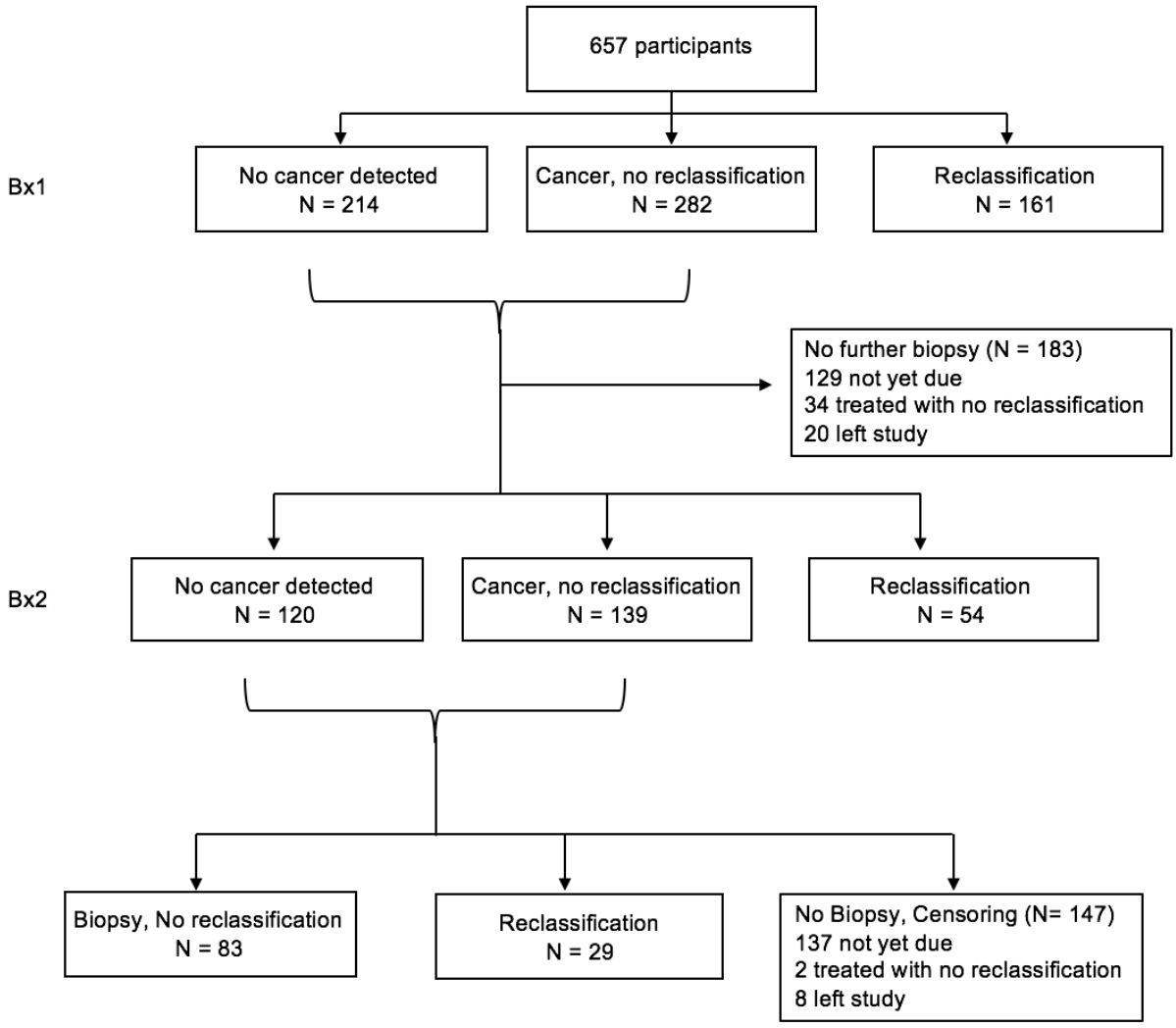
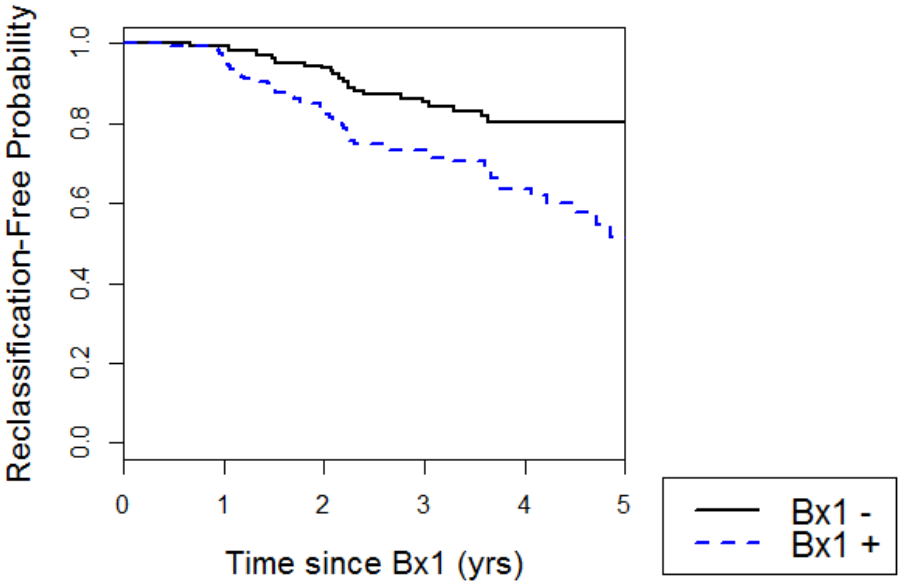


Figure 2

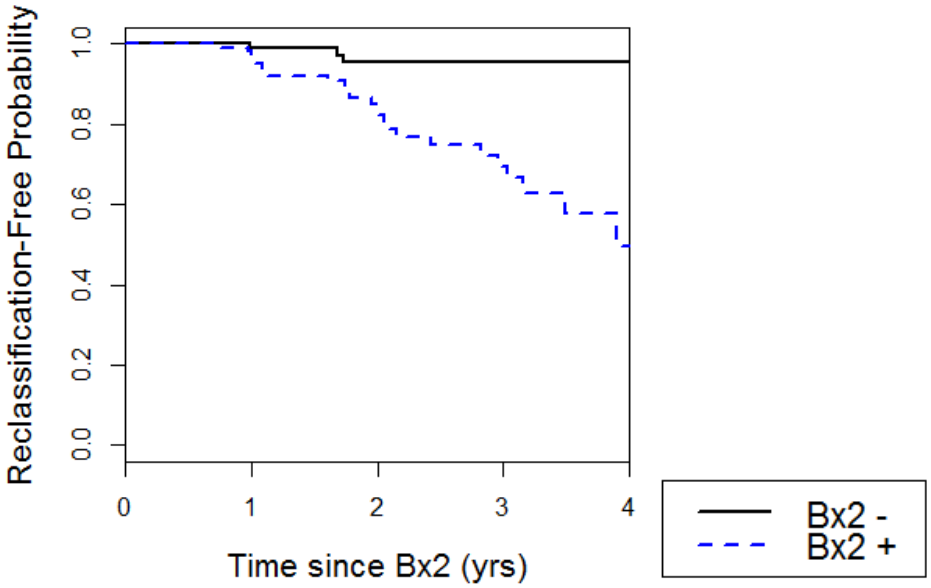


Log-rank test $p < 0.001$

Number at risk:

	Time since Bx1 (years)					
	0	**1**	**2**	**3**	**4**	**5**
Bx1 -	214	176	132	90	40	18
Bx1 +	282	196	132	82	41	12

Figure 3



Log-rank test $p < 0.001$

Number at risk:

	Time since Bx2 (years)				
	0	**1**	**2**	**3**	**4**
Bx2 -	120	88	55	26	10
Bx2 +	139	98	57	26	5